Narrowband Array Signal Processing Using Time-Frequency Distributions

Vom Fachbereich
Elektrotechnik und Informationstechnik
der Technischen Universität Darmstadt
zur Erlangung des Grades eines
Doktor-Ingenieurs (Dr.-Ing.)
genehmigte

Dissertationsschrift

von

Luke A. Cirillo

geboren am 17. Juli 1979 in Perth, Australien

Darmstadt 2007
Hochschulkennziffer D-17

Vorsitzender:	Prof. Dr.-Ing. Peter Meißner
Erstreferent:	Prof. Dr.-Ing. Abdelhak M. Zoubir
Korreferent:	Prof. Dr. Moeness G. Amin

Tag der Einreichung:	05.02.2007
Tag der mündlichen Prüfung:	10.05.2007

For my family

Abstract

In many engineering applications where sensor arrays are employed, such as radar, sonar, telecommunications, speech processing and medical imaging, the signals observed are often nonstationary. This dissertation addresses particular problems in narrowband array signal processing for nonstationary signals. By making use of spatial time-frequency distributions, one is able to effectively exploit the nonstationary nature of the source signals, with one caveat: the time-frequency localization of the sources should be known *a priori*, or must needs be estimated. The task of determining the time-frequency localization properties of signals from noise-contaminated sensor array measurements, is composed in this work as problems of ‘point selection’ and ‘signature estimation’.

A ‘point selection’ scheme for automatically determining the time-frequency locations at which spatial time-frequency distribution matrices exhibit underlying diagonal or off-diagonal structure is proposed, based on multiple hypothesis testing. The tendered method is used to achieve blind source separation of nonstationary signals via joint diagonalization and joint off-diagonalization of a set of spatial time-frequency distribution matrices.

Toward the goal of ‘signature estimation’, a computationally attractive implementation of a time-frequency Hough transform is proposed. Statistical analysis of the method is conducted to determine the achievable estimation accuracy. The proposed approach is applied to direction-of-arrival estimation, based on the averaging of spatial time-frequency distribution matrices.

The problem of micro-Doppler signature estimation is also examined. Micro-Doppler signatures arise, for example in radar, due to the vibrational or rotational motion of targets. The aforementioned signature estimation approach is shown to yield biased estimates of the micro-Doppler amplitude, and a bias correction procedure is given. The methods developed here are applied to data from a radar experiment for validation of the theoretical ideas.

Near-field parameter estimation is also considered. When sources are in the near-field of an array, it is possible to perform passive localization in both range and direction. The use of spatial time-frequency distributions for near-field localization is investigated. A means of distinguishing between the time-frequency representations of far- and near-field sources is also proposed. Data from an experimental radar system is analyzed using the proposed techniques.

Zusammenfassung

In vielen Anwendungen der Signalverarbeitung, in denen Sensorgruppen zum Einsatz kommen, wie z.B. Radar, Sonar, Telekommunikation, Sprachsignalverarbeitung und bildgebende Verfahren der Medizintechnik, sind die beobachteten Signale nicht stationär. Diese Doktorarbeit behandelt verschiedene Probleme der Sensorgruppensignalverarbeitung für schmalbandige, nicht stationäre Signale. Unter Verwendung von räumlichen Zeit-Frequenz-Verteilungen ist es möglich die nichtstationären Eigenschaften der Quellensignale auszuwerten. Dabei muss die Zeit-Frequenz-Lokalisierung entweder vorher bekannt sein oder muss geschätzt werden. Die Lokalisierung von nichtstationären Signalen, die durch Sensorgruppen geschätzt wird, wird in dieser Arbeit als Punkt-Selektion und Signatur-Schätzung bezeichnet.

Ein Verfahren für die Punkt-Selektion, basierend auf einem multiplen Hypothesentest, wird vorgestellt, das zur automatischen Bestimmung der Zeit-Frequenz-Lokalisierung verwendet werden kann. Die Matrizen der räumlichen Zeit-Frequenz-Verteilungen haben eine diagonale oder nichtdiagonale Struktur. Das vorgeschlagene Verfahren wird verwendet, um blinde Quellentrennung von nichtstationären Signalen unter Verwendung der sogenannten “joint diagonalization” und “joint off-diagonalization” aus einer Menge von Matrizen der räumlichen Zeit-Frequenz Verteilungen zu erzielen.

Für die Signatur-Schätzung wird eine schnelle Implementierung der Zeit-Frequenz-Hough-Transformation vorgestellt. Diese Implementierung wird statistisch analysiert, um die erreichbare Genauigkeit des Schätzers zu bestimmen. Dieser Ansatz wird auf ein Problem der Richtungsschätzung angewandt, die auf der Mittelung von räumlichen Zeit-Frequenz-Verteilungsmatrizen basiert.

Weiterhin wird das Problem der Mikro-Doppler-Signatur-Schätzung untersucht. Radarsignale weisen beispielsweise Mikro-Doppler-Effekte auf, die durch eine Vibrations- oder Rotationsbewegung des Zielobjektes entstehen können. Die eben erwähnte Methode zur Signatur-Schätzung führt allerdings zu systematischen Fehlern der Amplitude des Mikro-Doppler-Effekts. Aus diesem Grund wird ein Verfahren vorgeschlagen, das den systematischen Fehler korrigiert. Um die theoretischen Ergebnisse zu evaluieren, werden die entwickelten Methoden auf echte Daten angewandt, die bei einem Radar-Experiment erzeugt wurden.

Des weiteren wird die Parameterschätzung für das Nahfeld behandelt. Wenn sich Quellen im Nahfeld einer Sensorgruppe befinden, ist es möglich eine passive Lokalisierung der Quellen zu erzielen. Dabei können Richtung und Abstand der Quellen geschätzt werden. Die Verwendung von räumlichen Zeit-Frequenz-Verteilungen für die Nahfeld-Lokalisierung wird untersucht. Außerdem wird eine Methode entwickelt, die es ermöglicht zwischen Quellen von Nah- und Fernfeld zu unterscheiden. Hierbei werden ebenfalls echte Daten von einem Radar-Experiment verwendet, um die vorgeschlagene Methode zu analysieren.

Contents

List of Figures

List of Tables

Abbreviations, Acronyms and Symbols

AWGN	Additive white Gaussian noise
AMUSE	Algorithm for multiple signal classification
BSS	Blind source separation
CRB	Cramer-Rao bound
cdf	Cumulative distribution function
DML	Deterministic maximum likelihood
DOA	Direction-or-arrival
EM	Electro-magnetic
FDR	False discovery rate
FF	Far-field
FM	Frequency modulation/modulated
FWE	Family-wise error
HAF	Higher-order ambiguity function
HOS	Higher-order statistics
ICA	Independent component analysis
IID	Independent identically distributed
IF	Instantaneous frequency
JBD	Joint block diagonalization
JD	Joint diagonalization
JD/OD	Joint diagonalization and off-diagonalization
JOD	Joint off-diagonalization
LFM	Linear frequency modulation
LOS	Line of sight
MHT	Multiple hypothesis test(ing)
ML	Maximum likelihood
NF	Near-field
NLL	Negative log-likelihood (function)

NLILS	Nonlinear instantaneous least-squares
NLLS	Nonlinear least-squares
PCA	Principle component analysis
PPT	Polynomial phase transform
pdf	Probability density function
PWHT	Pseudo Wigner-Hough transform
PWVD	Pseudo Wigner-Ville Distribution
QFM	Quadratic frequency modulation
RID	Reduced interference distributions
RMSE	Root mean-squared error
RoA	Region of attraction
SAD	Sensor-angle distribution
SAT	Single auto-term
SFM	Sinusoidal frequency modulation
SML	Stochastic maximum likelihood
SNR	Signal-to-noise ratio
SOBI	Second-order blind identification
SOS	Second-order statistics
SRB	Sequentially rejective Bonferroni (test)
STFD	Spatial time-frequency distribution
SVD	Singular-value decomposition
TITO	Two-input two-output
TF	Time-frequency
TFA	Time-frequency analysis
UCA	Uniform circular array
ULA	Uniform linear array
WHT	Wigner-Hough transform
WVD	Wigner-Ville distribution

$\lvert x\rvert$	Absolute value of x
$\approx$	Approximately equal to
$\lceil x \rceil$	"Ceiling" of x (smallest integer greater than x)
$\mathbb{C}$	The set of complex numbers
x^*	Complex conjugate of x
$\mathsf{Cov}\left[X, Y\right]$	Covariance of X and Y
$\mathsf{Cum}\left[X_1, \ldots, X_n\right]$	Cumulant of $X_1, \ldots, X_n$

$\mathcal{C}_{\boldsymbol{X}}\left[ijkl\right]$	Cumulant of elements $[X_k, X_j, X_k, X_l]$ of the random vector $\boldsymbol{X}$
$\mathsf{Diag}[\boldsymbol{M}]$	Diagonal elements of $\boldsymbol{M}$ as a vector
M_{ij}	Element at ith row and jth column of $\boldsymbol{M}$
$\hat{\theta}$	Estimate or estimator of θ
$\mathsf{E}\left[X\right]$	Expectation of X
$\lfloor x \rfloor$	"Floor" of x (largest integer smaller than x)
$\lVert\boldsymbol{M}\rVert$	Frobenius norm of $\boldsymbol{M}$
$\boldsymbol{M}^{\mathrm{H}}$	Hermitian (conjugate) transpose of $\boldsymbol{M}$
$\mathcal{H}\left[x(t)\right]$	Hilbert transform of $x(t)$
$\mathsf{Im}\left[x\right]$	Imaginary part of x
$\mathsf{Im}\left[x\right]$	Imaginary part of x
$\odot$	Hadamard (element-wise) matrix product
$\boldsymbol{M}^{\dagger}$	Moore-Penrose pseudo inverse of $\boldsymbol{M}$
$\mathsf{Off}\left[\boldsymbol{M}\right]$	"Off" operator (sum of the magnitude of off-diagonal elements of $\boldsymbol{M}$)
$O(x)$	Order notation (the same order as x)
$\mathsf{P}\left[A\right]$	Probability of A
$\mathbb{R}$	The set of real numbers
$\mathsf{Re}\left[x\right]$	Real part of x
$\mathsf{Tr}\left[\boldsymbol{M}\right]$	Trace of $\boldsymbol{M}$
$\boldsymbol{M}^{\mathrm{T}}$	Transpose of $\boldsymbol{M}$
$\mathsf{Var}\left[X\right]$	Variance of X
$\mathbb{Z}$	The set of integers

Publications

The following publications have been produced during the period of PhD candidacy.

Internationally Refereed Journal Articles

1. L. Cirillo, A. Zoubir and M. Amin. "Blind source separation in the time-frequency domain based on multiple hypothesis testing." *IEEE Transactions on Signal Processing*, (accepted).

2. L. Cirillo, A. Zoubir and M. Amin. "Estimation of FM parameters using a time-frequency Hough transform." *IEEE Transactions on Signal Processing*, (under review).

3. P. Heidenreich, L. Cirillo and A. Zoubir. "Direction finding and source number detection of FM signals using spatial time-frequency distributions." *Signal Processing*, (in preparation).

Internationally Refereed Conference Papers

1. L. Cirillo, A. Zoubir and M. Amin. "Estimation of near-field parameters using spatial time-frequency distributions." *Proceedings of the 32*$^{\text{nd}}$ *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Honolulu, Hawaii, April 2007.

2. P. Heidenreich, L. Cirillo and A. Zoubir. "Direction finding of nonstationary signals using spatial time-frequency distributions and morphological image processing." *Proceedings of the 32*$^{\text{nd}}$ *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Honolulu, Hawaii, April 2007.

3. C. Euler, L. Cirillo and A. Zoubir. “Performance comparison of blind source separation algorithms using spatial time-frequency distributions.” *Proceedings of the 9th IEEE International Symposium on Signal Processing and its Applications (ISSPA)*, Sharjah, U.A.E., February, 2007.

4. L. Cirillo, A. Zoubir and M. Amin. “Estimation of FM parameters using a time-frequency Hough transform.” *Proceedings of the 31st IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Toulouse, France, May, 2006.

5. L. Cirillo and A. Zoubir. “On blind separation of nonstationary signals.” *Proceedings of the 8th IEEE International Symposium on Signal Processing and its Applications (ISSPA)*, Sydney, Australia, August, 2005.

6. L. Cirillo, A. Zoubir and M. Amin. “Direction finding of nonstationary signals using a time-frequency Hough transform.” *Proceedings of the 30th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Philadelphia, USA, March, 2005.

7. L. Cirillo and A. Zoubir. “A bootstrap scheme for time-frequency auto-term selection in antenna arrays.” *Proceedings of the 29th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Montreal, Canada, May, 2004.

8. L. Cirillo and M. Amin. “Auto-term detection using time-frequency array processing.” *Proceedings of the 28th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Hong Kong, China, April, 2003.

9. L. Cirillo, A. Zoubir and A. Gershman. “Direction-of-arrival estimation for uncorrelated FM signals.” *Proceedings of the 2nd IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Rosslyn, VA USA, August, 2002.

10. L. Cirillo, A. Zoubir, N. Ma and M. Amin. “Automatic classification of auto- and cross-terms of time-frequency distributions in antenna arrays.” *Proceedings of the 27th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Orlando, FL USA, May, 2002.

11. L. Cirillo, A. Zoubir and M. Amin. “Selection of auto- and cross-terms for blind non-stationary source separation.” *Proceedings of the 2nd IEEE*

International Symposium on Signal Processing and Information Technology (ISSPIT), Cairo, Egypt, December, 2001.

Other publications

1. L. Cirillo. "Point selection for time-frequency array processing of non-stationary signals." *Proceedings of the 2*$^{\text{nd}}$ *ATcrc Telecommunications and Networking Conference and Workshop*, Fremantle, Australia, October, 2002.
2. L. Cirillo. "Blind source separation of FM signals." *Proceedings of the 3*$^{\text{rd}}$ *Inter University Postgraduate Electrical Engineering Symposium (IU-PEES)*, Rockingham, Australia, October, 2002.

Acknowledgments

First and foremost, I would like to thank my supervisor, Prof. Abdelhak Zoubir, for his support, encouragement and guidance throughout the entire course of my postgraduate studies. I would also like to thank my co-supervisor, Prof. Moeness Amin, whose valuable feedback and advice were always greatly appreciated.

A special thanks to those colleagues who gave me the chance to work with experimental data, including Dr. Gordon Frazer from the DSTO in Australia and Prof. Amin and his team at the Center for Advanced Communications at Villanova University in the U.S.A..

Over the course of this work I have had the pleasure of working with many different people in both Australia and Germany, who have helped and supported me in many different ways. To all those people, too numerous to name, I express my sincere gratitude. This undertaking would have not been possible without you.

Finally I would like to thank my family: To Mum, Dad, Nick, Melita and Juliette, for always being there for me and giving your unconditional love and support throughout my life. And to my wife Khishig and my daughter Sara, whose love, support, understanding and patience have been a constant blessing, motivation and inspiration for me.

Darmstadt, June 2007

Chapter 1

Introduction

In a wide range of engineering applications, a collection or "array" of measurement sensors is employed to solve the problem at hand. The type of sensors employed will depend on the given application, for example antennas measuring an electrical field are used in telecommunications and radar, while sensors for measuring pressure fluctuations are needed in acoustic applications such as sonar, ultrasound and speech processing. Regardless of the type of sensor employed, by using more than one sensor, one may acquire more information about the measured phenomena.

Typically, the placement of sensors in different physical locations is performed to exploit any spatial diversity in the signal being measured, and to potentially infer spatial characteristics about the underlying process. For example, in a radar or sonar application, one may wish to determine from which direction an echo is returning, and thus infer the position of a target. In speech processing, it may be desirable to extract the speech signal from a speaker standing in a known position, while suppressing any "noise" coming from other locations, in order to improve intelligibility of the speech in hands-free communications or improve the performance of a speech recognition program.

In many real-world problems, the characteristics of the signals acquired by the sensors are varying with time. This may be a property of the originating process, such as with speech, or due to the surrounding environment, as when the measurement system is in motion with respect to the source of interest. A familiar example is the changing pitch of an ambulance siren as it drives past an observer. Due to the motion of the ambulance, the siren initially sounds "higher" as it approaches, then "lower" as it is driving away. This phenomenon is known as the Doppler effect, named after J. Christian A. Doppler who first hypothesized on the nature of this occurrence, which arises in many situations.

On the one hand, estimation of parameters linked to the time-varying nature of a signal may be enhanced through the use of multiple sensors. This may be relevant, for example, if one wishes to infer the velocity of a moving target with a radar system. On the other hand, if the time-varying nature of the process is expected, this property may be used to enhance the estimation of spatial parameters related to the physical location of the signal source, for example the position of a target observed using radar. It is specifically the combined consideration of both the time-varying characteristics of a measured signal and the spatial information provided by an array of measurement sensors, which forms the cynosure of this dissertation.

A particular field of signal processing, motivated by the existence of *nonstationary* signals, i.e. signals with time-varying spectral characteristics, is that of joint time-frequency analysis (TFA) [48]. While the notions and tools of TFA have been developed over more than seventy years[1] the application of TFA directly to sensor array processing problems has only begun to receive great interest over the last ten years. In particular, the notion of a spatial time-frequency distribution (STFD), introduced by Belouchrani and Amin in [20], has emerged as a powerful means of exploiting nonstationarity for solving multi-sensor problems.

The technical aspects of this dissertation are primarily based on the use of STFDs to solve problems such as blind source separation and direction-of-arrival estimation of nonstationary sources. These approaches focus on the spatial information to be gained by using an array of sensors. Also considered is the use of TFA for estimation of parameters describing nonstationary signals, in particular frequency modulated signals, which arise in many applications. The remainder of this chapter is organized as follows: In the following section, the array processing model is introduced and the processing problems which are dealt with in this dissertation are defined. The research objectives of this work are stated in Section 1.2 and the original contributions made by the author are then summarised in Section 1.3. Finally, an overview of this dissertation is given in Section 1.4.

---

[1] For a more detailed discussion on the history and development of time-frequency analysis, the reader is referred to [48].

1.1 Array signal processing

1.1.1 The array processing model

In the section we develop the model used to describe the signals received at an array of sensors with distinct spatial locations, due to the emission or reflection of signal energy from certain sources. The physical environment is depicted in Figure 1.1 where the angles θ and ϕ are termed the azimuth and elevation angles of the source, respectively. In general, there will be a number of sources present, each with distinct spatial locations. We assume that the sources are points in space, from which the propagation of signal energy originates. This may be due to, for example, emission of electro-magnetic (EM) energy from a transmitter in a wireless communications system, reflection of EM energy from a target in a radar system, or reflection of acoustic energy in a sonar system.

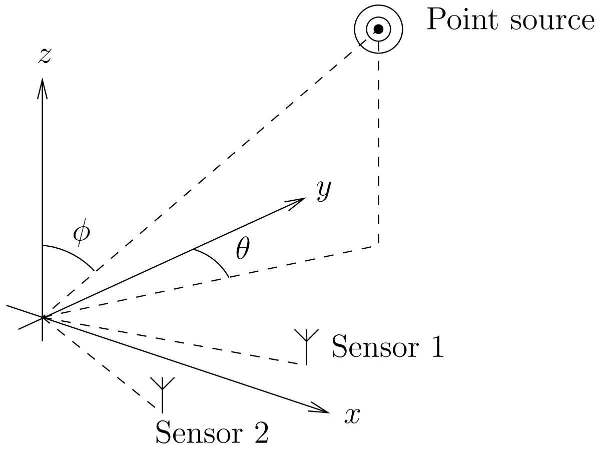


Figure 1.1: Sensor array environment containing two sensors and a point source.

The signal energy is described by a wave propagating radially outward from the source location, resulting in a spherical *wavefront* on which all points share the same phase. If the physical size of the sensor array is very small compared with the distance to the source, the wavefront may be considered as a plane across the array and the source is said to be in the *far-field*. On the other hand, if the sphericity of the wavefront must be taken into account, the sources are said to be in the *near-field* of the array.

The signal model developed in this section will be based on the far-field assumption. For simplicity, we shall also consider that the sources and array are coplanar, which means that the azimuth angle θ is the only relevant spatial parameter of a source. Furthermore, it is assumed that source signals are *narrowband* (have a small bandwidth) with respect to a known carrier frequency. These simplifying assumptions (which are consistent with applications such as wireless communications and radar) will be useful in the following development of the sensor array signal model.

We consider an array of m sensors with locations $\{\boldsymbol{r}_k\}_{k=1}^{m}$, where the first sensor is taken as a reference and assumed to reside at the spatial origin ($\boldsymbol{r}_1 = \mathbf{0}$). The source signal waveform with respect to time t, at a given point $\boldsymbol{r}$ in space, is denoted $S(\boldsymbol{r}, t)$. Under the narrowband signal representation, the signal waveform at the origin is

$$S(0, t) = s(t)\mathrm{e}^{\mathrm{j}\omega_c t}$$

where ω_c is the known carrier frequency and $s(t)$ is slowly time-varying with respect to the carrier term. We assume that the direction of propagation with respect to the origin is given by a vector[2] $\boldsymbol{r}_s$ having magnitude $\|\boldsymbol{r}_s\| = 1/c$, where c is the speed of propagation. Under the far-field assumption, the signal wavefront is planar and the waveform at sensor k is simply a time-shifted version of the waveform at the origin, as illustrated in Figure 1.2. The signal waveform at sensor k is therefore given by

$$S(\boldsymbol{r}_k, t) = s(t - \boldsymbol{r}_s \cdot \boldsymbol{r}_k)\mathrm{e}^{\mathrm{j}\omega_c (t - \boldsymbol{r}_s \cdot \boldsymbol{r}_k)}.$$

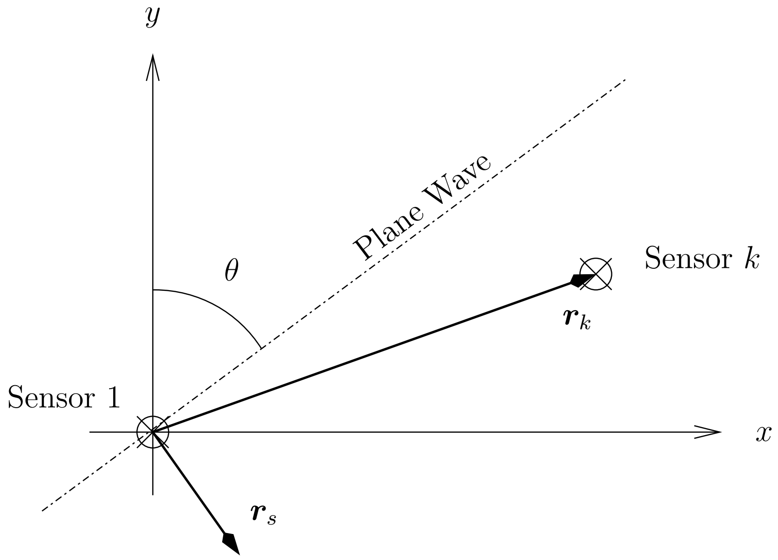


Figure 1.2: A plane wave incident on a sensor array.

If the signal $s(t)$ has bandwidth ω_B then approximation

$$s(t - \boldsymbol{r}_s \cdot \boldsymbol{r}_k) \approx s(t) \tag{1.1}$$

---

[2] $\boldsymbol{r}_s$ has been referred to as the slowness vector in the literature.

holds provided $|\boldsymbol{r}_s \cdot \boldsymbol{r}_k| << 2\pi/\omega_B$. Given the carrier has wavelength $\lambda = 2\pi c/\omega_c$ and since $|\boldsymbol{r}_s \cdot \boldsymbol{r}_k| \leq \|\boldsymbol{r}_s\|\|\boldsymbol{r}_k\|$, the requirement of Equation (1.1) can be expressed as

$$\varrho << \omega_c/\omega_B, \tag{1.2}$$

where $\varrho = \|\boldsymbol{r}_k\|/\lambda$. When $\boldsymbol{r}_k$ is the location of the sensor furthest from the reference sensor (origin), ϱ corresponds to the array size in wavelengths, also termed the *array aperture*. Assuming that Equation (1.2) is satisfied, then the following approximation holds:

$$S(\boldsymbol{r}_k, t) \approx s(t)\mathrm{e}^{\mathrm{j}\omega_c t}\mathrm{e}^{-j\boldsymbol{r}_k \cdot \boldsymbol{k}_s}, \tag{1.3}$$

where $\boldsymbol{k}_s = \omega_c \boldsymbol{r}_s$ is called the wave vector of the source and has magnitude $\|\boldsymbol{k}_s\| = 2\pi/\lambda = k$, called the wave-number. In the following, it is assumed that the inverse relative bandwidth of the source signal is much greater than the array aperture (i.e. Equation (1.2) holds).

We assume that the sensor measurement of the space-time field is linear and time-invariant and that the sensor output is demodulated with the known carrier frequency. The baseband output of sensor k is therefore given by

$$x_k(t) = g_k(\theta)\mathrm{e}^{-j\boldsymbol{r}_k \cdot \boldsymbol{k}_s} s(t), \tag{1.4}$$

where $g_k(\theta)$ denotes the gain of sensor k in the direction θ. By organizing the outputs of each sensor into a vector;

$$\boldsymbol{x}(t) = [x_1(t), \ldots, x_m(t)]^{\mathrm{T}},$$

and defining the vector

$$\boldsymbol{a}(\theta) = [g_1(\theta)\mathrm{e}^{-j\boldsymbol{r}_1 \cdot \boldsymbol{k}_s}, \ldots, g_m(\theta)\mathrm{e}^{-j\boldsymbol{r}_m \cdot \boldsymbol{k}_s}]^{\mathrm{T}},$$

one may represent the complete array output in vector form as

$$\boldsymbol{x}(t) = \boldsymbol{a}(\theta)s(t). \tag{1.5}$$

The vector $\boldsymbol{a}(\theta)$ describes the response of the array to a source at azimuth θ and is often referred to in the literature as the *steering vector*. The set of steering vectors formed as θ is varied across the parameter range of interest is termed the *array manifold*. For the array manifold to be known, the physical locations of the sensors, $\{\boldsymbol{r}_k\}_{k=1}^m$, as well as the sensor gain functions $\{g_k(\theta)\}_{k=1}^m$, must be known. In practice this requires a finely calibrated array of sensors.

In the case of multiple sources, we defined the kth source signal waveform and azimuth angle by $s_k(t)$ and θ_k respectively, for $k = 1, \ldots, d$, where d is the number of sources present. Since the sensors are assumed to act as linear systems, we may express the array output vector as a sum of the array responses to each individual source;

$$\boldsymbol{x}(t) = \sum_{i=k}^{d} \boldsymbol{a}(\theta_k) s_k(t) = \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{s}(t), \tag{1.6}$$

where

$$\begin{aligned} \boldsymbol{s}(t) &= [s_1(t), \ldots, s_d(t)]^{\mathrm{T}}, \\ \boldsymbol{\theta} &= [\theta_1, \ldots, \theta_d]^{\mathrm{T}}, \\ \text{and } \boldsymbol{A}(\boldsymbol{\theta}) &= [\boldsymbol{a}(\theta_1), \ldots, \boldsymbol{a}(\theta_d)]^{\mathrm{T}}. \end{aligned}$$

In order to account for inaccuracies in modeling the array response, for various sources of noise and interference, etc., an additive noise term $\boldsymbol{V}(t)$ is also included in the array model;

$$\boldsymbol{X}(t) = \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{s}(t) + \boldsymbol{V}(t). \tag{1.7}$$

Due to the stochastic nature of $\boldsymbol{V}(t)$ one may consider $\boldsymbol{X}(t)$ to be a vector or multi-channel continuous-time random process. We note that, depending on the context, it may also be appropriate to model the source signals as stochastic processes. This assumption is made, depending on the application.

1.1.2 Array processing problems

We assume that the demodulated sensor outputs are sampled at certain time instants $\{t_n\}_{n=1}^{N-1}$, yielding a set of N vector "snapshots" of the array output. For the sake of notational convenience, we shall denote the snapshot at time t_n by $\boldsymbol{X}(n)$. We shall also assume that a constant sampling interval Δ is used such that $\{t_n = n\Delta + t_0\}_{n=0}^{N-1}$, where t_0 denotes the time at which the first snapshot is taken. The data model, based on Equation (1.7), is

$$\boldsymbol{X}(n) = \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{s}(n) + \boldsymbol{V}(n); \quad n = 0, \ldots, N-1, \tag{1.8}$$

where the set of snapshots, $\{\boldsymbol{X}(n)\}_{n=0}^{N-1}$, is modelled as a multi-channel discrete-time random process, whose characteristics are specified by a multivariate probability density function.

Depending on what *a priori* information is available, one may define different problems to be solved given the data model in Equation (1.8). In this dissertation, we shall consider the following problems:

- Estimation of the source azimuth angles, $\boldsymbol{\theta}$, when the array manifold is known.
- Estimation of the mixing matrix, $\boldsymbol{A}$, or its inverse, when the array manifold is unknown.

The first problem listed above is referred to as *direction-of-arrival* (DOA) *estimation* or *direction finding.* The second problem is known as *blind source separation* (BSS). In order to solve the stated problems, certain assumptions on the signal and noise are made, which are stated below.

Assumption 1 (White noise) *We assume that the noise is a circular complex process satisfying*

$$\mathsf{E}\left[\boldsymbol{V}(n+\kappa)\boldsymbol{V}^{\mathrm{H}}(n)\right] = \sigma_v^2\delta(\kappa)\boldsymbol{I} \quad \textit{and} \quad \mathsf{E}\left[\boldsymbol{V}(n+\kappa)\boldsymbol{V}^{\mathrm{T}}(n)\right] = \mathbf{0}\ \forall\kappa. \tag{1.9}$$

Equation (1.9) implies that the noise is zero-mean and both spatially and temporally white.

Assumption 2 (Uncorrelated sources) *If a stochastic model for the sources is used, we assume that*

$$\mathsf{E}\left[\boldsymbol{S}(n+\kappa)\boldsymbol{S}^{\mathrm{H}}(n)\right] = \mathsf{Diag}[[P_1(\kappa),\ldots,P_d(\kappa)]^{\mathrm{T}}]. \tag{1.10}$$

If the source waveforms are considered to be deterministic function of time, then we assume that

$$\lim_{N\to\infty} N^{-1}\sum_{n=0}^{N-1} \boldsymbol{s}(n+\kappa)\boldsymbol{s}^{\mathrm{H}}(n) = \mathsf{Diag}[[P_1(\kappa),\ldots,P_d(\kappa)]^{\mathrm{T}}]. \tag{1.11}$$

The assumptions in Equation (1.10) or (1.11) imply that the sources are spatially uncorrelated, though they may be temporally correlated if $P_l(\kappa) \neq 0$ for $\kappa > 0$. The value $P_l(0)$ is termed the power of source l, for $l = 1,\ldots,d$.

1.2 Research objectives

The primary research objective is the *development of advanced array signal processing algorithms for nonstationary signals.* This objective is to be achieved in the context of *blind source separation* and *direction-of-arrival estimation.*

1.3 Summary of original contributions

The original contributions of this dissertation are summarised below.

1. *Automatic time-frequency point selection*
   The development of procedures for automatically determining the time-frequency locations where spatial time-frequency distribution matrices have desired underlying diagonal or off-diagonal structure.

2. *Blind source separation*
   The application of automatic time-frequency point selection procedures to the problems of blind source separation via joint diagonalization and joint off-diagonalization algorithms.

3. *Time-frequency signature estimation*
   The development of a computationally efficient means of estimating time-frequency signatures using a time-frequency Hough transform.

4. *Direction-of-arrival estimation*
   The application of the time-frequency Hough transform to the problem of direction-of-arrival estimation, including the under-determined case where the number of signals exceeds the number of sensors.

5. *Micro-Doppler signature estimation*
   The application of the time-frequency Hough transform to estimation of micro-Doppler signatures arising from vibrational or rotational motion of targets observed by a passive radar array.

6. *Near-field parameter estimation*
   The application of spatial time-frequency distributions to the near-field estimation problem and the development of a method for discriminating the time-frequency representations of far- and near-field sources from an observed mixture.

1.4 Overview of the dissertation

An overview of this dissertation is given below. The structure of the dissertation is also depicted graphically in Figure 1.3. It should be noted that Chapters 2 and 3 are principally review chapters, while the major contributions of this work are contained in Chapters 4 and 5. Chapters 6 and 7 present results obtained using experimental data from different radar systems.

Chapter 2: A summary of conventional array processing methods is given. Both problems of blind source separation and direction-of-arrival estimation are considered. The methods reviewed in this chapter are quite general in nature and do not deal specifically with the case of nonstationary signals.

Chapter 3: The application of TFA to array processing problems is reviewed. A brief introduction to TFA is followed by a discussion on STFDs. Approaches to blind source separation and direction-of-arrival estimation using STFD matrices are reviewed. This chapter provides the most essential background material for topics covered later in the dissertation.

Chapter 4: The topic of time-frequency 'point selection' is discussed in detail and a solution based on multiple hypothesis testing is presented. The proposed point selection approach is applied to the problem of blind source separation of nonstationary signals.

Chapter 5: The topic of time-frequency 'signature estimation' is discussed in detail and a solution based on a time-frequency Hough transform is presented. The proposed signature estimation approach is applied to the problem of direction-of-arrival estimation of nonstationary signals.

Chapter 6: The specific application of micro-Doppler parameter estimation is considered. The time-frequency signature estimation procedure developed in Chapter 5 is applied to both simulated and experimental data. The estimator bias is also considered.

Chapter 7: In contrast to the preceding work, in this chapter it is considered that the sources may be in the near-field of the sensor array. Estimation of near-field parameters using STFDs is investigated. A method for discriminating between the time-frequency representations of near- and far-field sources is proposed.

Chapter 8: The conclusions of this work are presented and possible directions for future research are discussed.

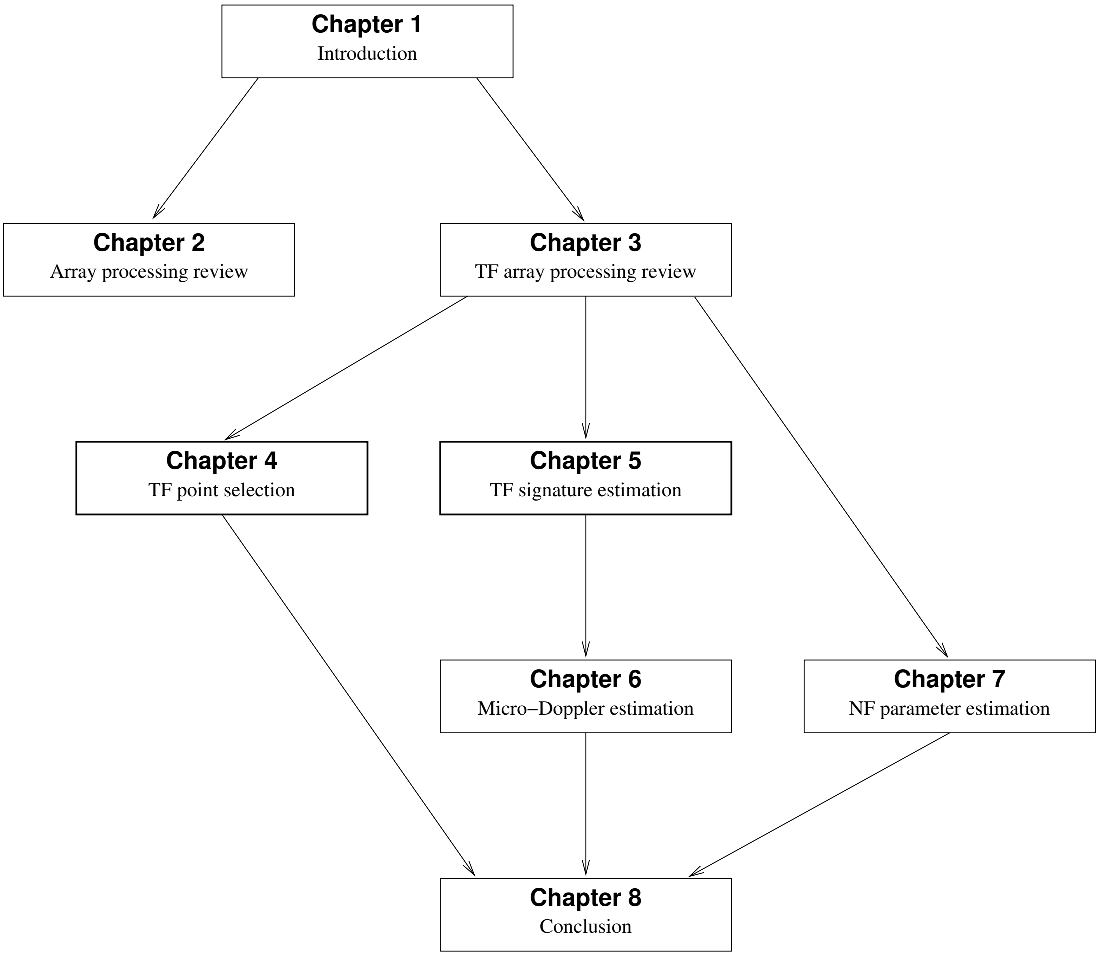


Figure 1.3: The structure of the dissertation.

Chapter 2

Conventional Array Processing Methods

The purpose of this chapter is to provide a review of "conventional" methods for blind source separation (in Section 2.1) and direction-of-arrival estimation (in Section 2.2). By "conventional," we refer to approaches which either assume (require) stationarity or do not exploit nonstationarity if it is present. Due to the huge body of literature available on array processing topics, the review given here is not meant to be exhaustive by any means, rather to give the "flavour" of the field by concentrating on the more well-cited and seminal works. We note that the methods presented herein are confined to the narrowband model introduced in Chapter 1 and do not consider the problem of convolutive or wideband models, which are beyond the scope of this dissertation.

2.1 Blind source separation

The field of BSS is relatively new, though it has developed rapidly over the past two decades. During this period, considerable developments have been made and many different approaches to solving the problem have been considered. In this section, a brief overview of some of the most important and fundamental approaches to BSS is provided, which will serve as an appropriate background for material covered in the following chapters.

2.1.1 Problem

The data model from Section 1.1.1 is repeated here for convenience, where the dependence on the source DOAs, $\boldsymbol{\theta}$, is not assumed to be known for BSS:

$$\boldsymbol{X}(n) = \boldsymbol{A}\boldsymbol{S}(n) + \boldsymbol{V}(n); \quad n = 0, \dots, N-1. \tag{2.1}$$

We note that in Equation (2.1) the sources are assumed to be stochastic. The sampled sensor outputs, $\boldsymbol{X}(n)$ are modelled by a discrete-time random process. The problem is to estimate the source signal waveforms modeled by $\boldsymbol{S}(n)$. There are, however, inherent indeterminacies in the model of Equation (2.1), which mean that (without further assumptions) the sources may be estimated only up to an arbitrary scaling and permutation. The indeterminacy is illustrated by writing Equation (2.1) in the form:

$$\boldsymbol{X}(n) = \sum_{k=1}^{d} \left(\frac{\boldsymbol{a}_{\sigma(k)}}{\alpha_k} \right) \big(\alpha_k S_{\sigma(k)}(n) \big) + \boldsymbol{V}(n); \quad n = 0, \dots, N-1, \tag{2.2}$$

where, for $k = 1 \dots, d$; $\alpha_k \in \mathbb{C}$ is a scalar, $\boldsymbol{a}_k$ denotes the kth column of $\boldsymbol{A}$ and $\sigma(k)$ denotes the kth element of some permutation on $\{1, \dots, d\}$. Let us then define a permutation matrix $\boldsymbol{P} \in \mathbb{R}^{d\times d}$ containing a single unit entry on every row and column, and a diagonal scaling matrix $\boldsymbol{D} \in \mathbb{C}^{d\times d}$.

The goal of the blind source separation, is to estimate a matrix $\boldsymbol{B}$, such that $\boldsymbol{B}\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}$.

2.1.2 Identifiability conditions

Conditions which guarantee that the matrix $\boldsymbol{B}$ may be estimated from the model in Equation (2.1) are termed *identifiability conditions* and generally require certain assumptions on the sources to be made. The two most common assumptions made are that the sources are *mutually independent* or *mutually uncorrelated.* Some identifiability conditions based on these two assumptions are outlined here:

2.1.2.1 Mutually independent sources

(C1) Let the sources be independently, identically distributed (IID) with respect to time and have joint pdf satisfying

$$f_{\boldsymbol{S}}(\boldsymbol{s}) = \prod_{k=1}^{d} f_{S_k}(s_k). \tag{2.3}$$

The noise-free version of the model in Equation (2.1) is identifiable provided that at most one source is Gaussian [49].

(C2) Let the sources be IID with respect to time and have joint pdf satisfying Equation (2.3). The model in Equation (2.1) is identifiable provided that [111]:

$$\frac{\mathsf{E}\,[S_k^4]}{\mathsf{E}\,[S_k^2]^2} \neq \frac{\mathsf{E}\,[S_l^4]}{\mathsf{E}\,[S_l^2]^2}; \quad \forall k \neq l. \tag{2.4}$$

The condition (C2) actually implies (C1) since the fourth order moments of Gaussian sources are specified by the second-order moments. These conditions are based on the assumption of IID sources and therefore do not make any use of (i.e. do not assume) any time-domain structure in the sources. The separation is therefore purely spatial in nature (i.e. based on the linear independence of the columns of $\boldsymbol{A}$).

2.1.2.2 Mutually uncorrelated sources

Let us first define the average power of a complex stochastic sequence $\{S(n)\}_{n=0}^{N-1}$ as:

$$r_S \triangleq N^{-1} \sum_{n=0}^{N-1} \mathsf{E}\left[|S(n)|^2\right]. \tag{2.5}$$

(C3) Let the sources be wide-sense stationary and temporally colored such that:

$$\boldsymbol{R}_{\boldsymbol{SS}}(\kappa) \triangleq \mathsf{E}\left[\boldsymbol{S}(n+\kappa)\boldsymbol{S}^{\mathrm{H}}(n)\right] = \mathsf{Diag}[[r_{S_1}(\kappa), \ldots, r_{S_d}(\kappa)]], \tag{2.6}$$

where $\{r_{S_k}(\kappa) \in \mathbb{C}\}_{k=1}^{d}$ are the auto-correlation sequences of the sources. The model in Equation (2.1) is identifiable provided that $\exists \kappa_0 \neq 0$ such that [111]

$$\frac{r_{S_k}(\kappa_0)}{r_{S_k}} \neq \frac{r_{S_l}(\kappa_0)}{r_{S_l}}; \quad \forall k \neq l. \tag{2.7}$$

(C4) Let the sources be nonstationary and temporally colored such that:

$$\boldsymbol{R}_{\boldsymbol{SS}}(n,\kappa) \triangleq \mathsf{E}\left[\boldsymbol{S}(n+\kappa)\boldsymbol{S}^{\mathrm{H}}(n)\right] = \mathsf{Diag}[[r_{S_1}(n,\kappa), \ldots, r_{S_d}(n,\kappa)]], \tag{2.8}$$

where $\{r_{S_k}(n,\kappa) \in \mathbb{C}\}_{k=1}^{d}$ are the 'local' auto-correlation sequences of the sources. The model in Equation (2.1) is identifiable provided that $\exists(n_0,\kappa_0)$ where $\kappa_0 \neq 0$ such that [20]

$$\frac{r_{S_k}(n_0,\kappa_0)}{r_{S_k}} \neq \frac{r_{S_l}(n_0,\kappa_0)}{r_{S_l}}; \quad \forall k \neq l. \tag{2.9}$$

The conditions (C3) and (C4) allow for separation of sources using only second-order statistics, provided there is some time-domain structure which can be exploited (i.e. coloration). These conditions could also be equivalently cast in the frequency domain due to the uniqueness of the Fourier transform; conditions (C3) and (C4) require the sources to have unique normalized spectra.

2.1.3 Orthogonalization

In many BSS algorithms, a whiteness constraint is placed on the data which allows separation via an orthogonal transformation. Implementation of the whiteness constraint usually requires a whitening transform to be first applied to the data, before separation can be achieved. Calculation of the whitening transform for the model in Equation (2.1) is outlined here, based on the derivation in [111].

A singular value decomposition of the mixing matrix $\boldsymbol{A} \in \mathbb{C}^{m \times d}$ yields

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V} \tag{2.10}$$

where $\boldsymbol{U} \in \mathbb{C}^{m \times d}$ is termed the signal subspace and satisfies $\boldsymbol{U}^{\mathrm{H}}\boldsymbol{U} = \boldsymbol{I}$, $\boldsymbol{\Lambda} \in \mathbb{C}^{d \times d}$ is a positive definite diagonal matrix and $\boldsymbol{V} \in \mathbb{C}^{d \times d}$ is orthogonal. The whitening transform matrix is defined as

$$\boldsymbol{W} \triangleq \boldsymbol{\Lambda}^{-1}\boldsymbol{U}^{\mathrm{H}}, \tag{2.11}$$

which means that

$$\boldsymbol{W}\boldsymbol{A} = \boldsymbol{\Lambda}^{-1}\boldsymbol{U}^{\mathrm{H}}\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V} = \boldsymbol{V},$$

which is orthogonal. Applying the transformation $\boldsymbol{W}$ to the data results in an orthogonal transformation of the sources, which, given one of the conditions (C1)-(C4) holds, implies that the signal part of the observations is spatially 'white'.

Calculation of $\boldsymbol{W}$ from Equation (2.11) depends on the unknown mixing system, $\boldsymbol{A}$. One can, however, estimate the whitening transformation from the observations. Consider the covariance matrix estimator:

$$\hat{\boldsymbol{R}}_{\boldsymbol{XX}} \triangleq N^{-1} \sum_{n=0}^{N-1} \boldsymbol{X}(n)\boldsymbol{X}(n)^{\mathrm{H}}. \tag{2.12}$$

Applying the expectation operator to Equation (2.12) yields

$$\boldsymbol{R}_{\boldsymbol{XX}} \triangleq \mathsf{E}\left[\hat{\boldsymbol{R}}_{\boldsymbol{XX}}\right] = \boldsymbol{A}\boldsymbol{R}_{\boldsymbol{SS}}\boldsymbol{A}^{\mathrm{H}} + \sigma_V^2 \boldsymbol{I} \tag{2.13}$$

where

$$\boldsymbol{R}_{\boldsymbol{SS}} \triangleq N^{-1} \sum_{n=0}^{N-1} \mathsf{E}\left[\boldsymbol{S}(n)\boldsymbol{S}^{\mathrm{H}}(n)\right].$$

We note that, under assumptions (C1)-(C4), the matrix $\boldsymbol{R}_{\boldsymbol{SS}}$ is diagonal. Based on the indeterminacy illustrated in Equation (2.2), one can assume, without loss of generality, that $\boldsymbol{R}_{\boldsymbol{SS}} = \boldsymbol{I}$, where the dynamic range of the sources is accounted for in the matrix $\boldsymbol{A}$. An SVD of the matrix $\boldsymbol{R}_{\boldsymbol{XX}}$ therefore yields

$$\begin{aligned} \boldsymbol{R}_{\boldsymbol{XX}} &= \boldsymbol{U}_X \boldsymbol{\Lambda}_X \boldsymbol{U}_X^{\mathrm{H}} \\ &= [\boldsymbol{U}\boldsymbol{U}_V](\boldsymbol{\Lambda}^2 + \sigma_V^2 \boldsymbol{I})[\boldsymbol{U}\boldsymbol{U}_V]^{\mathrm{H}} \end{aligned}$$

where $\boldsymbol{U}$ and $\boldsymbol{\Lambda}$ are the same matrices as in Equation (2.10).

A given estimate, $\hat{\boldsymbol{R}}_{\boldsymbol{xx}}$, of $\boldsymbol{R}_{\boldsymbol{XX}}$ is obtained from observations $\{\boldsymbol{x}(n)\}_{n=0}^{N-1}$ by replacing $\boldsymbol{X}(n)$ with $\boldsymbol{x}(n)$ in Equation (2.12). Based on the *sample* covariance matrix $\hat{\boldsymbol{R}}_{\boldsymbol{xx}}$ one may estimate the whitening transform as outlines in Table 2.1. The whitening transform has the advantage of reducing the computational complexity of the BSS process, as the dimension of the data is reduced from m to d. Also, estimation of the unknown orthogonal mixing matrix $\boldsymbol{V}$ is easier than directly trying to estimate the non-orthogonal mixing matrix $\boldsymbol{A}$. The disadvantage of whitening is a reduced interference-to-signal ratio (ISR), which becomes bounded from below [33].

1. Compute the sample covariance matrix:

$$\hat{\boldsymbol{R}}_{\boldsymbol{xx}} = N^{-1} \sum_{n=0}^{N-1} \boldsymbol{x}(n)\boldsymbol{x}^{\mathrm{H}}(n)$$

2. Compute the eigen-values of $\hat{\boldsymbol{R}}_{\boldsymbol{xx}}$, denoted $\lambda_1 \leq \cdots \leq \lambda_m$, and the corresponding eigen-vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m$.

3. Estimate the noise variance:

$$\hat{\sigma}_V^2 = (m-d)^{-1} \sum_{k=1}^{m-d} \lambda_k$$

4. Construct the diagonal scaling matrix:

$$\hat{\boldsymbol{\Lambda}} = \mathsf{Diag}[[\sqrt{\lambda_{m-d+1} - \hat{\sigma}_v^2}, \ldots, \sqrt{\lambda_m - \hat{\sigma}_v^2}]]$$

5. Construct the unitary matrix:

$$\hat{\boldsymbol{U}} = [\boldsymbol{u}_{m-d+1}, \ldots, \boldsymbol{u}_m]$$

6. Form an estimate of the whitening transform:

$$\hat{\boldsymbol{W}} = \hat{\boldsymbol{\Lambda}}^{-1}\hat{\boldsymbol{U}}.$$

Table 2.1: Estimation of the whitening transform.

2.1.4 Independent component analysis

Separation of sources based on the assumption of mutual independence was first discussed around 1985 by Herault and Jutten [69]. Because of similarities with principle component analysis (PCA), the problem was named independent component analysis (ICA), a term which is now prevalent in the BSS literature. The goal of ICA is to find an appropriate unmixing matrix, $\boldsymbol{B}$, such that $\boldsymbol{BX}(n)$ yields components which are statistically mutually independent.

The initial work of Jutten and Herault was motivated from a biological nervous system which transmits mixtures of information regarding, e.g. joint movement, over nervous fibers. The central nervous system is then able to separate the relevant information from the received mixtures. Since neither the source signals sent nor the mixing system are observable, the extraction of the sources must be done in a 'blind' manner, assuming only independence of the sources.

The neural model for source separation can be viewed as a recursive linear adaptive filter, which was discussed in detail in [80] for the two input two output (TITO) case. The separation structure is illustrated in Figure 2.1, where the coefficients, $c_{12}(n)$ and $c_{21}(n)$ are adapted until independence of the filter outputs is achieved. With respect to the problem as defined in Section 2.1.1, the separation matrix is $\boldsymbol{B}(n) = (\boldsymbol{I} + \boldsymbol{C}(n))^{-1}$ where $\boldsymbol{C}$ has zero entries on the diagonal, and the off-diagonal entries correspond to the filter feedback co-efficients.

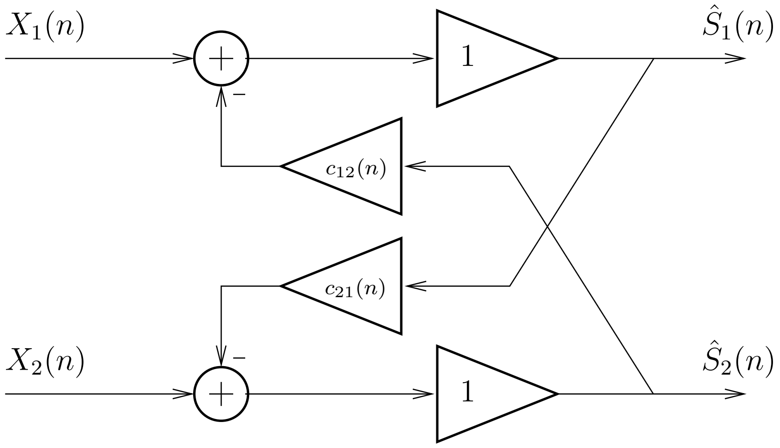


Figure 2.1: A neural network structure for BSS in the TITO case.

An adaptive algorithm for updating the filter coefficients of Figure 2.1 based on the gradient method was shown to lead to uncorrelated output signals in [80]. In order to achieve independent outputs, the following adaptation rule was proposed:

$$c_{kl}(n) = c_{kl}(n-1) + af(\hat{S}_k(n))g(\hat{S}_l(n)); \quad n = 1, \ldots, N-1, \tag{2.14}$$

where a is a positive adaptation gain, and $f(\cdot)$ and $g(\cdot)$ are two different nonlinear functions. Convergence based on Equation (2.14) implies statistical independence of the outputs provided the probability densities of the sources are even functions. One advantage of this adaptive approach is that the separation remains successful in time-varying mixtures provided the adaptation rate is faster than the evolution of the mixing system. One major disadvantage is that convergence and stability are not guaranteed and can depend greatly on the initialization of the filter [61,80].

We note that the method of Jutten and Herault makes implicit use of higher-order statistics in Equation (2.14), through the nonlinear functions $f(\cdot)$ and $g(\cdot)$, though no theoretical basis for the choice of the most appropriate functions is discussed. A more formal approach to designing ICA algorithms was proposed by Comon in [49] based on *contrast functions* of the probability distribution of the data. A summary which ties together several popular ICA algorithms through the use of contrast functions can be found in [35, 36]. Following the notational convention in [35], we shall denote a function, f, of a random vector, $\boldsymbol{X}$, as $f(\boldsymbol{X})$ and a function of the *distribution* of $\boldsymbol{X}$ as $f[\boldsymbol{X}]$.

In the following it is assumed that the observations are noise-free, i.e. $\{\boldsymbol{X}(n) = \boldsymbol{A}\boldsymbol{S}(n)\}_{n=0}^{N-1}$ in Equation (2.1), and the data satisfies condition (C1). A contrast function for BSS is denoted $\phi[\cdot]$ and is defined in such a way that $\phi[\boldsymbol{C}\boldsymbol{S}] \geq \phi[\boldsymbol{S}]$ with equality iff $\boldsymbol{C} = \boldsymbol{D}\boldsymbol{P}$, where $\boldsymbol{D}$ and $\boldsymbol{P}$ are the scale and permutation matrices defined in Section 2.1.1. Mixing of the sources increases the contrast function and minimization of the contrast function leads to separation of the sources. An algorithm for BSS is therefore defined by finding an appropriate contrast function and applying a minimization technique.

Maximum likelihood

BSS based on the maximum likelihood (ML) principle has been considered in [3, 64, 96]. The contrast function corresponding to the ML solution was shown in [34] to correspond to the Kullback-Leibler divergence between the distributions of the separated and true source signals. It was also shown in [34] that the ML contrast corresponds to the same contrast derived according to the infomax principle in [15]. The Kullback-Leibler divergence between the distributions of two random vectors $\boldsymbol{Y}$ and $\boldsymbol{S}$, is defined as

$$\mathcal{K}[\boldsymbol{Y}, \boldsymbol{S}] \triangleq \int f_{\boldsymbol{Y}}(\boldsymbol{\alpha}) \log\left(\frac{f_{\boldsymbol{Y}}(\boldsymbol{\alpha})}{f_{\boldsymbol{S}}(\boldsymbol{\alpha})}\right) d\boldsymbol{\alpha}, \tag{2.15}$$

where $f_{\boldsymbol{Y}}$ and $f_{\boldsymbol{S}}$ denote the probability density functions of $\boldsymbol{Y}$ and $\boldsymbol{S}$ respectively. $K[\cdot,\cdot]$ is considered a measure of how different two distributions are from one another. Defining $\boldsymbol{Y}(n) = \boldsymbol{B}\boldsymbol{X}(n)$ to be the separated sources, the maximum likelihood contrast function is shown in [34] to be asymptotically equivalent (as $N \to \infty$) to:

$$\phi_{ML}[\boldsymbol{Y}] = \mathcal{K}[\boldsymbol{Y}, \boldsymbol{S}], \tag{2.16}$$

where $\boldsymbol{Y}$ and $\boldsymbol{S}$ represent random vectors having the same distributions and $\boldsymbol{Y}(n)$ and $\boldsymbol{S}(n)$ respectively, for all $n = 0, \ldots, N-1$ (the sources are assumed IID with respect to n).

Minimization of Equation (2.16) with respect to $\boldsymbol{B}$ has been proposed using gradient based techniques [2,37]. The gradient of the Kullback-Leibler divergence between $\boldsymbol{Y}$ and $\boldsymbol{S}$ is given by [35]:

$$\nabla\phi_{ML}[\boldsymbol{Y}] = \nabla\mathcal{K}[\boldsymbol{Y}, \boldsymbol{S}] = \mathsf{E}\left[\boldsymbol{H}(\boldsymbol{Y})\right], \tag{2.17}$$

where $\boldsymbol{H}$ is the vector-to-matrix mapping:

$$\boldsymbol{H}(\boldsymbol{y}) = \boldsymbol{\varphi}(\boldsymbol{y})\boldsymbol{y}^{\mathrm{H}} - \boldsymbol{I} \tag{2.18}$$

and $\boldsymbol{\varphi}(\boldsymbol{y}) = \left[\varphi_1(y_1), \ldots, \varphi_d(y_d)\right]^{\mathrm{T}}$ is a vector of score-functions, defined in terms of the pdf of the sources according to

$$\varphi_k(\alpha) = \frac{\frac{\partial}{\partial\alpha} f_{S_k}(\alpha)}{f_{S_k}(\alpha)}; \quad k = 1, \ldots, d. \tag{2.19}$$

Given observations $\{\boldsymbol{x}(n)\}_{n=0}^{N-1}$ of the random process $\boldsymbol{X}(n)$, an on-line implementation [3] of a relative gradient descent algorithm for minimization of Equation (2.16) is given in Table 2.2, and an off-line implementation [96] in Table 2.3, based on the definition of $\boldsymbol{H}$ in Equation (2.18).

The same procedures given in Tables 2.2 and 2.3 may be implemented with a whiteness constraint [37] by modifying the gradient function according to:

$$\boldsymbol{H}(\boldsymbol{y}) = \boldsymbol{y}\boldsymbol{y}^{\mathrm{H}} - \boldsymbol{I} + \boldsymbol{\varphi}(\boldsymbol{y})\boldsymbol{y}^{\mathrm{H}} - \boldsymbol{y}^{\mathrm{H}}\boldsymbol{\varphi}(\boldsymbol{y}). \tag{2.20}$$

It is discussed in [36] that the whitening constraint provides some practical advantages, by allowing a two-stage implementation of the source separation process. In the first stage, a whitening transform is estimated (as outlined in Section 2.1.3), which is applied to the data. The optimisation of the contrast function is then with respect to an orthogonal matrix, which may be implemented efficiently by the Jacobi technique [38].

1. Initialize the separation matrix: $\boldsymbol{B}(0) = \boldsymbol{I}$.
2. Compute the initial estimate of the sources: $\boldsymbol{y}(0) = \boldsymbol{B}(0)\boldsymbol{x}(0)$.
3. For $n = 1, 2, 3, \ldots, N-1$:
   - Update the separation matrix according to:

$$\boldsymbol{B}(n) = (\boldsymbol{I} - \mu(n)\boldsymbol{H}(\boldsymbol{y}(n-1)))\boldsymbol{B}(n-1) \tag{2.21}$$

   - Update the estimate of the sources:$\boldsymbol{y}(n) = \boldsymbol{B}(n)\boldsymbol{x}(n)$

where $\{\mu(n) \in \mathbb{R}^+\}_{n=1}^{N-1}$ is a scalar sequence of learning steps.

Table 2.2: On-line implementation of the maximum-likelihood BSS algorithm.

1. Initialize the output: $\boldsymbol{y}(n) = \boldsymbol{x}(n)$, for $n = 0, \ldots, N-1$.
2. Estimate the average field: $\bar{\boldsymbol{H}} = N^{-1}\sum_{n=0}^{N-1} \boldsymbol{H}(\boldsymbol{y}(n))$.
3. Update the output:

$$\boldsymbol{y}(n) = (\boldsymbol{I} - \mu\bar{\boldsymbol{H}})\boldsymbol{y}(n); \quad n = 0, \ldots, N-1. \tag{2.22}$$

If $\bar{\boldsymbol{H}}$ is small, stop, else repeat from 2.

Table 2.3: Off-line implementation of the maximum-likelihood BSS algorithm.

Mutual information

In the maximum likelihood approach to BSS, the distribution of the sources is assumed known (or a good guess is available). If the distribution is unknown, the Kullback-Leibler divergence should be minimized with respect to both the separating matrix $\boldsymbol{B}$ and the source distribution $f_{\boldsymbol{S}}$. For a random vector $\boldsymbol{S}$ of independent entries, the Kullback-Leibler divergence between the separation output and the sources can be expressed as

$$\mathcal{K}[\boldsymbol{Y}, \boldsymbol{S}] = \mathcal{K}[\boldsymbol{Y}, \tilde{\boldsymbol{Y}}] + \mathcal{K}[\tilde{\boldsymbol{Y}}, \boldsymbol{S}], \tag{2.23}$$

where $\tilde{\boldsymbol{Y}}$ is a vector of independent entries, each having distribution given by the marginal distribution of the corresponding element of $\boldsymbol{Y}$. The divergence is therefore minimized by setting $\boldsymbol{S} = \tilde{\boldsymbol{Y}}$ which leads to the contrast function:

$$\phi_{MI}[\boldsymbol{Y}] = \mathcal{K}[\boldsymbol{Y}, \tilde{\boldsymbol{Y}}] = \min_{f_{\boldsymbol{S}}} \mathcal{K}[\boldsymbol{Y}, \boldsymbol{S}], \tag{2.24}$$

which is known as the mutual information contrast function.

The use of mutual information to define a source separation contrast was first proposed by Comon in [49]. Unfortunately, the contrast Equation (2.24) is not easily minimized using a gradient approach, as the gradient does not reduce to the expectation of a simple function of $\boldsymbol{Y}$, as is the case for maximum likelihood. Proposed approaches based on mutual information include the use of kernel estimates for the marginal distributions of $\boldsymbol{Y}$ as in [95] or parametric estimates of these distributions [86, 92]. Comon proposed the use of higher-order statistics in [49] to approximate Equation (2.24), based on an Edgeworth expansion.

Use of higher-order statistics

A number of approximations to the ML solution of the BSS problem have been proposed using higher-order statistics. An in-depth review and analysis of such techniques, in terms of the accuracy and computational efficiency of certain implementations, is given in [36]. The higher-order approximations are conveniently expressed in terms of cumulants. Typically, the second and fourth order cumulants are used, though in the following, only fourth-order cumulants shall be required due to the orthogonality constraint. Given real variables $X_1, \ldots, X_4$ and defining $\bar{X}_k \triangleq X_k - \mathsf{E}[X_k]$ for $k = 1, \ldots, 4$, the fourth-order cumulant is defined as:

$$\begin{aligned}\mathsf{Cum}[X_1, X_2, X_3, X_4] &= \mathsf{E}\left[\bar{X}_1\bar{X}_2\bar{X}_3\bar{X}_4\right] - \mathsf{E}\left[\bar{X}_1\bar{X}_2\right]\mathsf{E}\left[\bar{X}_3\bar{X}_4\right] \\ &\quad -\mathsf{E}\left[\bar{X}_1\bar{X}_3\right]\mathsf{E}\left[\bar{X}_2\bar{X}_4\right] - \mathsf{E}\left[\bar{X}_1\bar{X}_4\right]\mathsf{E}\left[\bar{X}_2\bar{X}_3\right].\end{aligned}$$

For a vector of complex random variables $\boldsymbol{X} = [X_1, \ldots, X_m]^{\mathrm{T}}$ the cumulants of elements of $\boldsymbol{X}$ shall be denoted $\mathcal{C}_{klmn}[\boldsymbol{X}] \triangleq \mathsf{Cum}[X_k, X_l^*, X_m, X_n^*]$.

In [49], Comon considers a fourth-order Edgeworth expansion of the pdf and a random variable about its best Gaussian approximation. The approximation of the mutual information contrast function for BSS under a whiteness constraint and assuming a symmetric distribution is given by:

$$\phi_{ICA}[\boldsymbol{Y}] = \sum_{klmn \neq kkkk} |\mathcal{C}_{klmn}[\boldsymbol{Y}]|^2. \tag{2.25}$$

An algorithm for minimization of Equation (2.25) for real signals is proposed by Comon based on Givens rotations. It is noted in [38], however, that the Givens angles can not be computed in closed form for the complex case.

In [38] a smaller subset of cross-cumulants was considered in forming the orthogonal contrast[1]:

$$\phi_{JADE}[\boldsymbol{Y}] = \sum_{klmn \neq kkmn} |\mathcal{C}_{klmn}[\boldsymbol{Y}]|^2, \tag{2.26}$$

which was motivated by the fact that Equation (2.26) can be used as a joint diagonalization criterion, allowing a computationally efficient implementation of the minimization of $\phi_{JADE}[\boldsymbol{Y}]$. Details of a Jacobi-like algorithm for the minimization of Equation (2.26) are given in [38], which apply to the complex case.

2.1.5 Separation using second-order statistics

The BSS methods discussed in the previous section generally make the assumption that the observed data is IID. In this respect, no time-domain structure is present which could be exploited. However, if the signals of interest have some coloration, i.e. there is some correlation between the signal at different points in time, then this information may also be exploited for source separation. Perhaps the simplest method for separation using second-order statistics is the Algorithm for Multiple Unknown Signals Extraction (AMUSE) [111].

The AMUSE algorithm involves computing a whitening transform (see Section 2.1.3), which is applied to the observed data to achieve an orthogonal transformation of the sources. The unknown orthogonal transformation is then determined by an eigen-decomposition of the whitened data covariance matrix, which is in turn used to estimate the mixing matrix or the source signals. This algorithm is summarised in Table 2.4. The main problem with this approach, is the selection of an appropriate time-lag at which to compute the covariance matrix. If the time-lag is not chosen correctly, the covariance matrix will not have distinct eigen-values and the separation procedure will fail.

One possible extension to the AMUSE algorithm, is to jointly diagonalize a number of covariance matrices to determine the unitary factor $\boldsymbol{V}$ in Table 2.4, which can increase robustness to the problem of degenerate eigen-values and improve statistical performance. It was proposed in [84] to perform the joint diagonalization of the covariance matrix at lag $\kappa = 0$ and at another value of $\kappa > 0$. We note that in this work, the noise-free version of Equation (2.1) was assumed. The estimate, $\hat{\boldsymbol{A}}$, of the mixing matrix is obtained by solving the equations

$$\hat{\boldsymbol{R}}_{\boldsymbol{xx}}(0) = \hat{\boldsymbol{A}}\Lambda_0\hat{\boldsymbol{A}}^{\mathrm{H}} \tag{2.27}$$

$$\hat{\boldsymbol{R}}_{\boldsymbol{xx}}(\kappa) = \hat{\boldsymbol{A}}\Lambda_\kappa\hat{\boldsymbol{A}}^{\mathrm{H}}, \tag{2.28}$$

---

[1] The acronym JADE stands for joint approximate diagonalization of eigen-matrices

1. Estimate the whitening transform, $\boldsymbol{W}$, according to the procedure in Table 2.1. Obtain the whitened data samples
$$\boldsymbol{z}(n) = \hat{\boldsymbol{W}}\boldsymbol{x}(n); \quad n = 0, \ldots, N-1.$$
2. Select a lag, $\kappa > 0$, and calculate the matrix:
$$\hat{\boldsymbol{R}}_{\boldsymbol{zz}}(\kappa) = (N-\kappa)^{-1} \sum_{n=0}^{N-\kappa-1} \boldsymbol{z}(n+\kappa)\boldsymbol{z}^{\mathrm{H}}(n),$$
such that $(\hat{\boldsymbol{R}}_{\boldsymbol{zz}}(\kappa) + \hat{\boldsymbol{R}}^{\mathrm{H}}_{\boldsymbol{zz}}(\kappa))/2$ has distinct eigen-values.
3. Calculate the matrix, $\boldsymbol{V}$, of eigen-vectors of $(\hat{\boldsymbol{R}}_{\boldsymbol{zz}}(\kappa) + \hat{\boldsymbol{R}}^{\mathrm{H}}_{\boldsymbol{zz}}(\kappa))/2$.
4. Estimate the mixing matrix, $\boldsymbol{A}$, according to: $\hat{\boldsymbol{A}} = \hat{\boldsymbol{W}}^{\dagger}\boldsymbol{V}$.
5. Estimate the signal waveforms, $\boldsymbol{s}(n)$, according to: $\hat{\boldsymbol{s}}(n) = \boldsymbol{V}^{\mathrm{H}}\boldsymbol{z}(n)$, for $n = 0, \ldots, N-1$.

Table 2.4: Summary of the AMUSE algorithm for blind source separation.

where Λ_0 and Λ_κ are unknown diagonal matrices. The combination of Equations (2.27) and (2.28) leads to the eigenvalue problem:

$$\left(\hat{\boldsymbol{R}}_{\boldsymbol{xx}}(0)\hat{\boldsymbol{R}}^{-1}_{\boldsymbol{xx}}(\kappa)\right)\hat{\boldsymbol{A}} = \hat{\boldsymbol{A}}\left(\Lambda_0\Lambda_\kappa^{-1}\right) \tag{2.29}$$

which can be solved using standard numerical linear algebra methods.

Further extending the AMUSE algorithm, is the proposed BSS technique of Belouchrani *et al* in [17]. In this approach, a number of covariance matrices are jointly diagonalized to determine the separating matrix. This is achieved via minimization of a joint diagonality (JD) criterion. The JD criterion given in [17] is expressed in terms of the matrix "off" operator, defined for a matrix $\boldsymbol{M} \in \mathbb{C}^{p\times p}$ as:

$$\mathsf{Off}\,[\boldsymbol{M}] \triangleq \sum_{1\leq k\neq l\leq n} |M_{kl}|^2. \tag{2.30}$$

Given a set $\mathcal{M} = \{\boldsymbol{M}_1, \ldots, \boldsymbol{M}_K\}$ of square matrices of dimension p, the JD criterion is defined as the non-negative function of a square matrix $\boldsymbol{V} \in \mathbb{C}^{p\times p}$ by:

$$\mathcal{C}(\mathcal{M}, \boldsymbol{V}) \triangleq \sum_{k=1}^{K} \mathsf{Off}\left[\boldsymbol{V}^{\mathrm{H}}\boldsymbol{M}_k\boldsymbol{V}\right]. \tag{2.31}$$

A unitary matrix which minimizes Equation (2.31) over the set of all unitary matrices is said to be a joint diagonalizer of $\mathcal{M}$. The uniqueness of joint diagonalization is discussed in [17]. The BSS algorithm based on Equation (2.31) is termed second order blind identification (SOBI) and is summarised in Table 2.5.

1. Estimate the whitening transform, $\boldsymbol{W}$, according to the procedure in Table 2.1. Obtain the whitened data samples
$$\boldsymbol{z}(n) = \hat{\boldsymbol{W}}\boldsymbol{x}(n); \quad n = 0, \ldots, N-1.$$
2. Select a set of K lags, $\{\kappa_1, \ldots, \kappa_K\}$ and calculate the set of sample covariance matrices:
$$\mathcal{M} \triangleq \{\hat{\boldsymbol{R}}_{\boldsymbol{zz}}(\kappa_k);\ k = 1, \ldots, K\}.$$
3. Find the unitary matrix, $\boldsymbol{V}$, which minimizes the JD criterion defined in Equation (2.31):
$$\boldsymbol{V} = \min_{\boldsymbol{U}} \mathcal{C}(\mathcal{M}, \boldsymbol{U}).$$
4. Estimate the mixing matrix, $\boldsymbol{A}$, according to: $\hat{\boldsymbol{A}} = \hat{\boldsymbol{W}}^{\dagger}\boldsymbol{V}$.
5. Estimate the signal waveforms, $\boldsymbol{s}(n)$, according to: $\hat{\boldsymbol{s}}(n) = \boldsymbol{V}^{\mathrm{H}}\boldsymbol{z}(n)$, for $n = 0, \ldots, N-1$.

Table 2.5: Summary of the SOBI algorithm for blind source separation.

We note that the SOBI algorithm makes use of the whitening transformation discussed in Section 2.1.3, therefore the matrix determined by the minimization of Equation (2.31) is orthogonal. This has the advantage that computationally efficient methods based on Jacobi rotations may be used in the minimization of the JD criterion [17]. However, as mentioned in Section 2.1.3, the BSS performance is limited through the use of a whitening transform. An approach based on *non-orthogonal* joint diagonalization has later been suggested by Yeredor [121] which minimizes a least-squares JD criterion without the orthogonal constraint, and offers improved BSS performance. Non-orthogonal JD shall be discussed further in the following chapter, in the context of BSS of nonstationary signals.

2.2 Direction-of-arrival estimation

The topic of DOA estimation has received considerable attention in the literature over the last three decades. A concise and approachable summary of the first two decades of research in this field may be found in [81], while a more recent and comprehensive treatment of the subject is complied in [112]. In this section, a brief overview of some of the most important and fundamental approaches to DOA estimation is provided, which will serve as appropriate background for material covered in the following chapters.

2.2.1 Problem

The signal model defined in Section 1.1.1 is repeated here for convenience:

$$\boldsymbol{X}(t) = \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{S}(t) + \boldsymbol{V}(t), \quad t \in \mathbb{R}, \tag{2.32}$$

where $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_d]^{\mathrm{T}}$ is the vector of source locations.

The goal of DOA estimation, is to estimate the vector $\boldsymbol{\theta}$*, given observations* $\{\boldsymbol{x}(n)\}_{n=1}^{N}$ *of the appropriately sampled random process* $\boldsymbol{X}(t)$*.*

2.2.2 Classical Beamforming

An early approach to direction-finding with sensor arrays, involved “looking” or “steering” the array toward all possible directions of interest successively, and calculating the received power in each direction. Signal directions are then said to be those at which a high received power is obtained. This approach has also been termed spectral-based [81], as one effectively searches through a spectrum-like function (of power), taking the locations of the highest peaks to be estimates of the direction.

The classical Beamformer is obtained by steering the array using a linear combination of the sensor outputs: $Y(t) = \boldsymbol{w}^{\mathrm{H}}\boldsymbol{X}(t)$, where $\boldsymbol{w}$ is denoted the weighting vector. Assuming that a finite amount of energy is available for beamforming ($\|w\|^2 = k$, $0 < k < \infty$), the maximum Beamformer output power is obtained according to

$$P_{max} = \max_{\|\boldsymbol{w}\|=k} \mathsf{E}\left[|Y(t)|^2\right] = \max_{\|\boldsymbol{w}\|=k} \boldsymbol{w}^{\mathrm{H}}\mathsf{E}\left[\boldsymbol{X}(t)\boldsymbol{X}^{\mathrm{H}}(t)\right]\boldsymbol{w}. \tag{2.33}$$

If a single source is present at spatial location θ_1, then according to the model in Equation (1.7), Equation (2.33) may be expressed as

$$P_{max} = \max_{\|\boldsymbol{w}\|=1} \left(\mathsf{E}\left[|S(t)|^2\right] |\boldsymbol{w}^{\mathrm{H}}\boldsymbol{a}(\theta_1)|^2 + \sigma_v^2\right), \tag{2.34}$$

where $k = 1$ is arbitrarily chosen without loss of generality. The solution to Equation (2.34) yields the Beamformer weighting vector: $\boldsymbol{w}_{BF} \triangleq \boldsymbol{a}(\theta_1)/\|\boldsymbol{a}(\theta_1)\|$. This means that the output power is maximized when the array is steered to the true source location.

Given measurements $\{\boldsymbol{x}(n)\}_{n=1}^{N}$ from the array sensors, the output power may be estimated as a function of the weighting vector $\boldsymbol{w}$:

$$P(\boldsymbol{w}) = N^{-1}\sum_{n=1}^{N}|y(n)|^2 = \boldsymbol{w}^{\mathrm{H}}\left(N^{-1}\sum_{n=1}^{N}\boldsymbol{x}(n)\boldsymbol{x}(n)^{\mathrm{H}}\right)\boldsymbol{w} = \boldsymbol{w}^{\mathrm{H}}\hat{\boldsymbol{R}}_{\boldsymbol{xx}}\boldsymbol{w} \qquad (2.35)$$

As the true source location θ_1 is unknown, one can substitute the Beamformer weighting vector into Equation (2.35) as a function of the spatial parameter θ, which leads to the classical Beamformer spectrum[2]:

$$P_{BF}(\theta) = \frac{\boldsymbol{a}^{\mathrm{H}}(\theta)\hat{\boldsymbol{R}}_{\boldsymbol{xx}}\boldsymbol{a}(\theta)}{\|\boldsymbol{a}(\theta)\|^2}. \qquad (2.36)$$

Since the output power is maximized for $\theta = \theta_1$, an estimate of the source location is obtained according to

$$\hat{\theta}_1 = \arg\max_{\theta} P_{BF}(\theta). \qquad (2.37)$$

In cases where more than one source is present, it is common practice to take the locations of the largest separated peaks of Equation (2.36) as estimates of the sources' locations.

It is interesting to note that, while the Beamformer has been derived in a somewhat *ad hoc* manner, the estimator corresponding to Equation (2.37) is actually optimal in the sense of achieving statistical efficiency, when only one source is present. However, the main limitation of the Beamformer is its ability to resolve multiple closely-spaced sources. The spatial resolution of the Beamformer is determined solely by the array geometry, hence, collecting a larger number of observations does not help in resolving closely-spaced sources. Other approaches which better exploit the model of Equation (2.32) and can obtain better estimation accuracy are discussed in the following sections.

2.2.3 The MUSIC estimator

One of the most well-known approaches for DOA estimation is the MUlitiple SIgnal Classification (MUSIC) algorithm [99]. The MUSIC algorithm exploits the spectral expansion of the array covariance matrix according to: $\boldsymbol{R}_{\boldsymbol{XX}} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{H}}$, where $\boldsymbol{U}$ is an othonormal matrix of Eigen-vectors and $\boldsymbol{\Lambda}$ is a diagonal matrix of Eigen-values of $\boldsymbol{R}_{\boldsymbol{XX}}$. Given the model Equation (2.32) and assuming that $\boldsymbol{R}_{\boldsymbol{SS}}$ has full rank and $\boldsymbol{V}(t)$ is white noise according to Assumption 1, one may separate the spectral decomposition of the array covariance matrix into signal

---

[2] Also referred to as the spatial spectrum.

and noise components:

$$\boldsymbol{R}_{\boldsymbol{XX}} = \boldsymbol{U}_s \boldsymbol{\Lambda}_s {\boldsymbol{U}_s}^{\mathrm{H}} + \boldsymbol{U}_n \boldsymbol{\Lambda}_n {\boldsymbol{U}_n}^{\mathrm{H}}, \tag{2.38}$$

where $\boldsymbol{\Lambda}_s$ corresponds to the Eigen-values of $\boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{R}_{\boldsymbol{SS}}\boldsymbol{A}^{\mathrm{H}}(\boldsymbol{\theta})$ and $\boldsymbol{U}_s$ spans the same column space as $\boldsymbol{A}(\boldsymbol{\theta})$. The entries of $\boldsymbol{\Lambda}_n$ are called noise Eigen-values and are equal to the noise power σ_V^2, while the columns of $\boldsymbol{U}_n$ span the null-space of $\boldsymbol{A}(\boldsymbol{\theta})$ and are called noise Eigen-vectors.

The MUSIC estimator exploits the orthogonality between the signal and noise subspaces, which implies that:

$$\boldsymbol{U}_n^{\mathrm{H}} \boldsymbol{a}(\theta) = 0 \quad \text{for} \quad \theta \in \{\theta_1, \ldots, \theta_d\}. \tag{2.39}$$

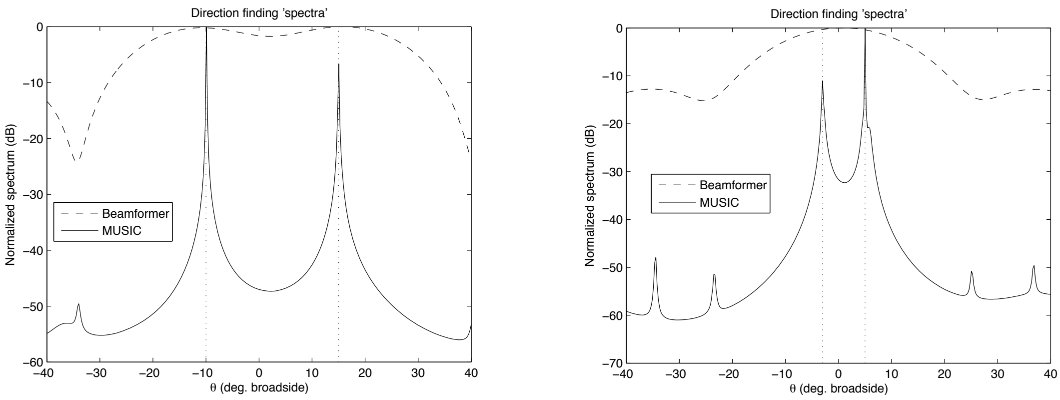
The MUSIC “spectrum” is therefore defined according to:

$$P_M(\theta) = \frac{1}{\|\hat{\boldsymbol{U}}_n^{\mathrm{H}} \boldsymbol{a}(\theta)\|^2}, \tag{2.40}$$

which will, theoretically, become infinitely large for $\theta \in \{\theta_1, \ldots, \theta_d\}$ as $\hat{\boldsymbol{U}}_n \to \boldsymbol{U}_n$. The estimate of $\boldsymbol{\theta}$ is therefore formed from the locations of the d largest peaks of $P_M(\theta)$. In practice $\boldsymbol{U}_n$ is estimated from the Eigen-vectors corresponding to the $m - d$ smallest Eigen-values of $\hat{\boldsymbol{R}}_{\boldsymbol{xx}}$.

An example of the MUSIC and the Beamformer spectrum is shown in Figure 2.2, for two incident plane waves on a uniform linear array of sensors. It can be seen that, for closely spaced signals, the MUSIC spectrum is able to resolve the two directions while the Beamformer fails. The dramatic resolution improvement over the Beamformer has lead to the wide popularity of MUSIC. In fact, assuming that the sensor data is accurately modelled, MUSIC can provide statistically consistent estimates for sufficiently large SNR and sample size [105]. It was also shown in [105] that for uncorrelated source signals, MUSIC corresponds to a large sample realization the ‘deterministic’ maximum likelihood estimator (which is discussed in the following section).

Although the MUSIC estimator has many attractive properties, it still demonstrates relatively poor spatial resolution for small sample sizes and at low SNR. The inability to resolve closely spaced sources is especially accentuated if the signals are highly correlated. In applications where coherent signals are present, the principle of the MUSIC algorithm given in Equation (2.39) no longer holds and the the MUSIC algorithm can not be applied. Given these limitations, there exist a large number of extensions and generalizations of MUSIC (see e.g. [9,30,57,58]), which aim to provide improved performance as well as more computationally efficient implementations.



(a) (b)

Figure 2.2: An example of the MUSIC and Beamformer spectra for the case of two plane waves incident on a uniform linear array of sensors. The true DOAs are indicated by dotted lines.

2.2.4 Maximum likelihood

Given an appropriate statistical model for the observations, the maximum likelihood (ML) approach may be applied to derive an estimator for the source directions [108]. The ML estimator for direction-finding provides improved accuracy w.r.t. the previously discussed methods, and can be applied in the cases of highly correlated and coherent signals. A detailed discussion of the ML principle can be found in most textbooks on estimation theory (see e.g. [83]), and will not be presented here. It is noted, however, that under certain regularity conditions, every consistent ML estimator is asymptotically efficient. Moreover, if the statistical model for the observations belongs to the exponential family (including the Gaussian distribution), all regularity conditions are fulfilled and every consistent ML estimator is asymptotically efficient.

The probability density function (pdf) of the random variables $\{\boldsymbol{X}(n)\}_{n=1}^{N}$ is denoted $f(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N; \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a vector of all unknown parameters (including the source locations $\boldsymbol{\theta}$). The ML estimator for $\boldsymbol{\theta}$ is given by:

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} f(\boldsymbol{X}(1), \ldots, \boldsymbol{X}(N); \boldsymbol{\theta}) \tag{2.41}$$

A particular estimate of $\boldsymbol{\theta}$ is obtained from a given set of observations by replacing $\{\boldsymbol{X}(n)\}_{n=1}^{N}$ with $\{\boldsymbol{x}(n)\}_{n=1}^{N}$ in Equation (2.41). The function to be maximized, $l(\boldsymbol{\theta}) = f(\boldsymbol{x}(1), \ldots, \boldsymbol{x}(N); \boldsymbol{\theta})$ is termed the *likelihood function.* It is also common practice to consider the *negative log likelihood function* (NLL), given by $\mathcal{L}(\boldsymbol{\theta}) \triangleq -\log l(\boldsymbol{\theta})$. As the logarithmic function is strictly monotonic, the minimum of $-\log l(\boldsymbol{\theta})$ and the maximum of $l(\boldsymbol{\theta})$ coincide, while the functional form of the log likelihood function is often much simpler (especially when dealing with exponential distributions) than that of the likelihood function.

Depending on which statistical model is assumed in Equation (2.32), the NLL can be derived and possibly "concentrated" with respect to the source locations parameters. The ML estimate of $\boldsymbol{\theta}$ is then obtained according to:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}). \tag{2.42}$$

Two different statistical assumptions are common in the literature and shall be reviewed here.

Deterministic maximum likelihood

The 'deterministic' maximum likelihood (DML) estimator is based on the following assumption:

Assumption 3 (Deterministic sources) *The source signals have unknown deterministic waveforms, $\boldsymbol{s}(t)$, while the noise is a white Gaussian process, $\boldsymbol{V}(t)$, satisfying Assumption 1.*

This assumption is intuitive in problems where the emitter signals arise from a specific (but possibly unknown) generating process, as opposed to the noise waveform which is assumed to arise from the combination of many independent sources. Based on Assumption 3 the NLL function (ignoring constants) is given by [31, 98]:

$$\mathcal{L}_{DML}(\boldsymbol{\theta}, \{\boldsymbol{s}(n)\}_{n=1}^{N}, \sigma_v^2) = m \log \sigma_v^2 + \frac{1}{\sigma_v^2 N} \sum_{n=1}^{N} \|\boldsymbol{x}(n) - \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{s}(n)\|^2, \tag{2.43}$$

where $\|\cdot\|$ denotes the Euclidean norm. Concentration of the NLL with respect to the source locations parameters [26, 116] yields:

$$\mathcal{L}_{DML}(\boldsymbol{\theta}) = \mathsf{Tr}\left[(\boldsymbol{I} - \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{A}^{\dagger}(\boldsymbol{\theta}))\hat{\boldsymbol{R}}_{\boldsymbol{xx}}\right], \tag{2.44}$$

where $(\cdot)^{\dagger}$ denotes the Moore-Penrose pseudo-inverse and

$$\hat{\boldsymbol{R}}_{\boldsymbol{xx}} = N^{-1} \sum_{n=1}^{N} \boldsymbol{x}(n)\boldsymbol{x}^{\mathrm{H}}(n)$$

is the sample covariance matrix. The DML estimate of the source locations is obtained via minimization of Equation (2.44) with respect to $\boldsymbol{\theta}$.

Stochastic maximum likelihood

After DML, the other popular ML formulation of the direction-finding problem is known as 'stochastic' maximum likelihood (SML) and is based on the following assumption:

Assumption 4 (Stochastic sources) *The source signals are modelled by a Gaussian processes, $\boldsymbol{S}(t)$, satisfying*

$$\mathsf{E}\left[\boldsymbol{S}(t+\tau)\boldsymbol{S}^{\mathrm{H}}(t)\right] = \boldsymbol{R_{SS}}(\tau) \quad \textit{and} \quad \mathsf{E}\left[\boldsymbol{S}(t+\tau)\boldsymbol{S}^{\mathrm{T}}(t)\right] = \boldsymbol{0}\ \forall\tau,$$

and the noise is a white Gaussian process, $\boldsymbol{V}(t)$, satisfying Assumption 1 and uncorrelated with $\boldsymbol{S}(t)$.

Modeling the source waveforms as stochastic processes may be appropriate, for example, where measurements are obtained by applying narrowband bandpass filters to wideband signals. The assumption that the sources are Gaussian leads to a tractable ML solution, though it should be noted that this approach is also applicable to non-Gaussian signals. In fact, the asymptotic estimation accuracy has been shown to depend only on the second-order statistics of the signal waveforms [90, 106].

The normalized NLL (ignoring constants) obtained using Assumption 4 is given by:

$$\mathcal{L}_{SML}(\boldsymbol{\theta}, \boldsymbol{R_{SS}}, \sigma_v^2) = \log|\boldsymbol{R_{XX}}| + \mathsf{Tr}\left[\boldsymbol{R}_{\boldsymbol{XX}}^{-1}\hat{\boldsymbol{R}}_{\boldsymbol{xx}}\right]. \tag{2.45}$$

It is also possible to concentrate Equation (2.45) with respect to the source location parameters, as for the DML case. A compact form for the concentrated NLL is:

$$\mathcal{L}_{SML}(\boldsymbol{\theta}) = \log|\boldsymbol{A}(\boldsymbol{\theta})\hat{\boldsymbol{R}}_{\boldsymbol{ss}}\boldsymbol{A}^{\mathrm{H}}(\boldsymbol{\theta}) + \hat{\sigma}_V^2\boldsymbol{I}|, \tag{2.46}$$

where the estimates $\hat{\sigma}_V^2$ and $\hat{\boldsymbol{R}}_{\boldsymbol{ss}}$ are given respectively by [27, 78]:

$$\begin{aligned}
\hat{\sigma}_V^2 &= \frac{1}{m-d}\mathsf{Tr}\left[\boldsymbol{I} - \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{A}^{\dagger}(\boldsymbol{\theta})\right] && (2.47)\\
\hat{\boldsymbol{R}}_{\boldsymbol{ss}} &= \boldsymbol{A}^{\dagger}(\boldsymbol{\theta})(\hat{\boldsymbol{R}}_{\boldsymbol{xx}} - \hat{\sigma}_V^2\boldsymbol{I})\boldsymbol{A}^{\dagger\mathrm{H}}(\boldsymbol{\theta}). && (2.48)
\end{aligned}$$

The SML estimate of the source locations is obtained via minimization of Equation (2.46) with respect to $\boldsymbol{\theta}$.

Implementation

The practical implementation of DOA estimates based on Equation (2.44) or Equation (2.46) can be a computationally demanding task. In general, the optimization problem of minimizing the NLL function requires a multi-dimensional search over a nonlinear and non-convex function. A typical approach is to obtain initial estimates of the DOAs using a less complex estimator, for example MUSIC. Given sufficiently accurate initial estimates, Gauss-Newton type algorithms may then be used to perform the minimization of the NLL function [32, 91, 104, 115]. However, if the initial estimates are poor, the search may not converge to the global minimum. Subspace approximations to the ML solution have also been proposed [108, 109, 114, 115], which yield the same statistical accuracy at reduced computational cost. For the specific case of uniform linear arrays (ULAs), the special structure of the matrix $\boldsymbol{A}$ allows for re-parametrization of the DML criterion in terms of the co-efficients of a polynomial having roots on the unit circle. This insight has lead to both iterative [29] and essentially closed-form solutions [109] for the DML estimator for ULAs.

2.3 Summary

In this chapter, a review of "conventional" narrowband array processing techniques has been provided. Two important signal processing problems have been considered: *blind source separation*; in which the structure of the mixing system is unknown, and *DOA estimation*; in which the mixing system can be parametrized in terms of the spatial locations of the sources.

In solving the BSS problem, the use of HOS may be seen as an approximation to the ML solution, when the sources are modelled as statistically independent IID random processes. Such approaches are appropriate for non-Gaussian sources. The use of SOS also provides an effective means for separating colored sources with different normalized spectra. A robust approach to implementing a number of BSS algorithms is based on the joint-diagonalization of a set of matrices composed with the HOS or SOS of the data. Using an orthogonal constraint for the JD provides more computationally attractive implementations at the cost of reduced BSS performance, in comparison with a non-orthogonal JD implementation.

DOA estimation methods allow one to infer the spatial location of the sources from the sensor array data, provided the array response can be appropriately parametrized. Estimation of the sources' locations can most easily be achieved by a linear combination of the sensor outputs, known as "beamforming". Beamforming does not, however, provide good resolution of closely-spaced sources. The subspace-based MUSIC algorithm is more complex than beamforming but can provide better resolution and consistent estimation. Accuracy is generally poor for highly correlated sources. ML estimators can provided consistent, efficient estimation at a large computational cost. The ML approach allows location estimation of highly correlated and coherent sources.

Chapter 3

Time-Frequency Array Processing Methods

The array processing methods discussed in Chapter 2 were not conceived with the application to nonstationary signals in mind. These"conventional" approaches exploit the spatial information or certain (stationary) signal structure for solving particular problems. As discussed in Chapter 1, there are many applications where sensor arrays are employed which deal with nonstationary signals. One well known and well developed approach to the analysis of nonstationarity is that of joint time-frequency analysis (TFA). In this chapter, a brief review of the important aspects of TFA is presented in Section 3.1. In Section 3.2 we discuss how TFA may be used to allow conventional array processing methods to exploit nonstationarity when it is present in the signals of interest. The chapter is summarized in Section 3.3 and the important problems to be addressed in the following chapters are outlined.

3.1 Time-frequency analysis

In the following, a review of basic time and frequency descriptions of signals is first provided, which paves the way for the use of joint time-frequency distributions, presented in Section 3.1.3. We will restrict our attention to the case of deterministic signals for the most part, while mentioning how the theory is extended to stochastic processes in Section 3.1.4.

3.1.1 Time and frequency descriptions of signals

A “signal” of interest usually arises from observing some physical phenomenon over time, e.g., an electromagnetic field, pressure, voltage, or sound waveform, denoted $s(t)$. It is assumed that the Fourier transform of $s(t)$ exists and is calculated according to:

$$S(\Omega) = \int_{\mathbb{R}} s(t) \mathrm{e}^{-\mathrm{j}\Omega t} dt, \tag{3.1}$$

where the signal is obtained from its Fourier transform using:

$$s(t) = \frac{1}{2\pi} \int_{\mathbb{R}} S(\Omega) \mathrm{e}^{\mathrm{j}\Omega t} d\Omega. \tag{3.2}$$

We note that Equation (3.1) provides an alternate description of the time-domain signal $s(t)$, in terms of frequency Ω. This description provides insights into the periodicities present in the signal which may not necessarily be obvious from its time-domain waveform.

In electrical engineering, one often wishes to know how much energy is required to produce (or is contained within) a given signal. In the case of electromagnetic theory, Poynting's theorem states that the electric (resp. magnetic) energy density is the absolute square of the electric (resp. magnetic) field. In general, whether ones talks about electromagnetic, voltage, sound or other signal waveforms, the signal energy is given by $|s(t)|^2$ and $|s(t)|^2 \Delta t$ is the amount of energy required to produce the signal in a small time interval of Δt, at time t. We shall denote the signal *energy density* herein as

$$D_{ss}(t) = |s(t)|^2 \tag{3.3}$$

and the *energy density spectrum* as

$$D_{ss}(\Omega) = \frac{1}{2\pi} |S(\Omega)|^2. \tag{3.4}$$

We note that the total energy contained within a signal is equivalently expressed in terms of time or frequency as:

$$E_s = \int_{\mathbb{R}} D_{ss}(t) dt \equiv \int_{\mathbb{R}} D_{ss}(\Omega) d\Omega. \tag{3.5}$$

A signal for which $E_s < \infty$, i.e. has finite total energy, is referred to as an *energy signal.*

While both the time and frequency descriptions of a signal provide an effective basis for signal analysis, they lack the ability to capture the dependency between time and frequency of a signal's energy distribution. For example, a large value of $D_{ss}(\Omega)$ at a given frequency $\Omega = \Omega_0$, tells us that there is significant signal energy at frequency Ω_0, but it does not indicate how this energy is distributed with respect to time. It may be spread out over the whole signal interval, contained in the first ten percent of the signal's duration, or be present intermittently. Clearly a joint time-frequency representation of the signal energy is needed in order to better analyze such signals.

3.1.2 Instantaneous frequency and the analytic signal

The concept of a signal having an instantaneous frequency (IF) is an intuitive idea which stems from observing many signals in nature, e.g. the changing color of the sky during a sunset, or the varying pitch from a musical instrument being played. The IF of a signal is therefore the frequency which is present at a particular instant in time. The most appropriate way to mathematically define the IF is not, however, obvious. A common approach in the literature, discussed in detail by Cohen in [48], is to define the IF as the derivative of the phase of the analytic signal.

The analytic signal is formed by considering only positive frequencies, i.e. the components $S(\Omega)$ for $\Omega > 0$. One forms the analytic signal, $z(t)$, corresponding to a signal $s(t)$, by using the Hilbert transform, denoted here by $\mathcal{H}[\cdot]$, according to:

$$z(t) = s(t) + \mathrm{j}\mathcal{H}[s(t)], \tag{3.6}$$

where

$$\mathcal{H}[s(t)] = \frac{1}{\pi}\int_{\mathbb{R}} \frac{s(t')}{t-t'}dt'. \tag{3.7}$$

The complex signal $z(t)$ may then be expressed in the polar form

$$z(t) = A(t)\mathrm{e}^{\mathrm{j}\psi(t)} \tag{3.8}$$

where $A(t)$ is termed the amplitude and $\psi(t)$ the phase. It was shown in [48] that the analytic procedure of Equation (3.6) effectively encodes the low frequency content of the signal into $A(t)$ and the high frequency content into $\mathrm{e}^{\mathrm{j}\psi(t)}$.

Given the signal as expressed in Equation (3.8), the IF, $\tilde{\omega}(t)$, is then defined as the derivative of the phase with respect to time:

$$\tilde{\omega}(t) = \frac{d}{dt}\psi(t). \tag{3.9}$$

The signal IF can be thought of as the conditional average of frequencies existing at a particular time. A related concept is the instantaneous bandwidth (IB), defined as:

$$B(t) = \left| \frac{\frac{d}{dt}A(t)}{A(t)} \right|. \tag{3.10}$$

The IB is a measure of the spread of frequencies at a particular time and can be thought of as the conditional standard deviation of frequencies about the IF. A more in depth discussion on the motivation for the definitions in Equations (3.9) and (3.10) can be found in [48].

3.1.3 Joint time-frequency descriptions of signals

As mentioned Section 3.1.1, the notion of a joint time-frequency description of signals is of great interest. The question of how to define the joint TF distribution of a signal has been the subject of signal processing research for many years. One of the most popular approaches, proposed by Cohen [47], is to define a distribution of the form

$$D_{zz}(t,\Omega) = \frac{1}{4\pi^2} \int_{\mathbb{R}^3} \phi(\theta,\tau) z(t' + \tau/2) z^*(t' - \tau/2) \mathrm{e}^{-\mathrm{j}\theta t - \mathrm{j}\Omega\tau + \mathrm{j}\theta t'} dt' d\tau d\theta, \tag{3.11}$$

where $\phi(\theta,\tau)$ is termed the kernel function [46]. The kernel functions for a number of TFDs are listed in Table 3.1. It is interesting that certain desirable properties of a TFD may be expressed in terms of conditions on the kernel function. Some of these properties are listed in Table 3.2. It is also interesting that certain properties listed in Table 3.2 may not be satisfied by some kernels which have nevertheless proved useful in signal analysis problems [14, 77].

Name	Kernel: $\phi(\theta,\tau)$
Wigner-Ville	1
Pseudo Wigner-Ville	$g(\tau)$
Margenau-Hill	$\cos(\theta\tau/2)$
Kirkwood-Rihaczek	$\mathrm{e}^{\mathrm{j}\theta\tau/2}$
Born-Jordan	$\frac{\sin(\theta\tau/2)}{\theta\tau/2}$
Page	$\mathrm{e}^{\mathrm{j}\theta\|\tau\|}$
Spectrogram	$\int h(u+\tau/2)h^*(u-\tau/2)\mathrm{e}^{-\mathrm{j}\theta u}du$
Zhao-Atlas-Marks	$g(\tau)\|\tau\|\frac{\sin(a\theta\tau)}{a\theta\tau}$
B-Distribution	$\|\tau\|^{\alpha}\frac{2^{2\alpha-1}}{\Gamma(2\alpha)}\Gamma(\alpha+j\pi\theta)\Gamma(\alpha+j\pi\theta)$

Table 3.1: Kernel functions of some well-known TFDs. The functions $g(\cdot)$ and $h(\cdot)$ denote windowing functions.

Property	Kernel condition	TFD satisfies
Time marginal	$\phi(\theta,0)=1$	$\int D_{xx}(t,\Omega)d\Omega = \lvert x(t)\rvert^2$
Frequency marginal	$\phi(0,\tau)=1$	$\int D_{xx}(t,\Omega)dt = \lvert X(\Omega)\rvert^2$
Total energy	$\phi(0,0)=1$	$\iint D_{xx}(t,\Omega)dt d\Omega = 1$
Reality	$\phi(\theta,\tau)=\phi^*(-\theta,-\tau)$	$\mathsf{Im}\,[D_{xx}(t,\Omega)] = 0$
Time-shift invariance	$\phi(\theta,\tau)$ independent of t	If $z(t)=x(t-t_0)$ then $D_{zz}(t,\Omega)=D_{xx}(t-t_0,\Omega)$
Frequency-shift invariance	$\phi(\theta,\tau)$ independent of Ω	If $z(t)=x(t)\mathrm{e}^{\mathrm{j}\Omega_0 t}$ then $D_{zz}(t,\Omega)=D_{xx}(t,\Omega-\Omega_0)$
Scaling invariance	$\phi(\theta,\tau)=\phi(\theta\tau)$	If $z(t)=\sqrt{a}x(at)$ then $D_{zz}(t,\Omega)=D_{xx}(at,\Omega/a)$
Weak finite time support	$\int \phi(\theta,\tau)\mathrm{e}^{-\mathrm{j}\theta t}d\theta = 0$ for $\lvert\tau\rvert \le 2\lvert t\rvert$	If $x(t)=0$ for $t\notin[t_1,t_2]$ then $D_{xx}(t,\Omega)=0$ for $t\notin[t_1,t_2]$
Weak finite frequency support	$\int \phi(\theta,\tau)\mathrm{e}^{-\mathrm{j}\tau\Omega}d\tau = 0$ for $\lvert\theta\rvert \le 2\lvert\Omega\rvert$	If $S(\Omega)=0$ for $\Omega\notin[\Omega_1,\Omega_2]$ then $D_{xx}(t,\Omega)=0$ for $\Omega\notin[\Omega_1,\Omega_2]$
Strong finite time support	$\int \phi(\theta,\tau)\mathrm{e}^{-\mathrm{j}\theta t}d\theta = 0$ for $\lvert\tau\rvert \ne 2\lvert t\rvert$	If $x(t)=0$ then $D_{xx}(t,\Omega)=0$
Strong finite frequency support	$\int \phi(\theta,\tau)\mathrm{e}^{-\mathrm{j}\tau\Omega}d\tau = 0$ for $\lvert\theta\rvert \ne 2\lvert\Omega\rvert$	If $S(\Omega)=0$ then $D_{xx}(t,\Omega)=0$
Covariance	$\left.\frac{\partial\phi(\theta,\tau)}{\partial\theta\partial\tau}\right\rvert_{\theta,\tau=0} = 0$	$\iint t\Omega D_{xx}(t,\Omega)dtd\Omega = \int t\psi'(t)\lvert x(t)\rvert^2 dt$
Instantaneous frequency	$\left.\frac{\partial\phi(\theta,\tau)}{\partial\theta\partial\tau}\right\rvert_{\tau=0} = 0$	$\psi'(t) = \frac{\int \Omega D_{xx}(t,\Omega)d\Omega}{\int D_{xx}(t,\Omega)d\Omega}$

Table 3.2: TFD properties and corresponding kernel conditions.

An alternative form for Equation (3.11) which will be used in the following, is obtained by defining the Fourier transform of the kernel function with respect to θ:

$$\varphi(t,\tau) = \frac{1}{2\pi}\int_{\mathbb{R}} \phi(\theta,\tau)\mathrm{e}^{-\mathrm{j}t\theta}d\theta. \tag{3.12}$$

One may then express the TFD given in Equation (3.11) in terms of the kernel function from Equation (3.12) as:

$$D_{zz}(t,\Omega) = \frac{1}{2\pi}\int_{\mathbb{R}^2} \varphi(t-t',\tau)z(t'+\tau/2)z^*(t'-\tau/2)\mathrm{e}^{-\mathrm{j}\Omega\tau}dt'd\tau. \tag{3.13}$$

It is also of interest to define a cross-TFD between two signals, which shall be referred to in the following chapters. The cross-TFD of signals $w(t)$ and $z(t)$ is defined as:

$$D_{wz}(t,\Omega) = \frac{1}{2\pi}\int_{\mathbb{R}^2} \varphi(t-t',\tau)w(t'+\tau/2)z^*(t'-\tau/2)\mathrm{e}^{-\mathrm{j}\Omega\tau}dt'd\tau. \tag{3.14}$$

We note that the TFD of a sum of a number of signals may be expressed in terms of the TFDs and cross-TFDs of the individual signals.

Multicomponent signals

In simple terms, "multicomponent" signals are those which exhibit multiple peaks or ridges in the time-frequency plane. A monocomponent signal is therefore one which exhibits only a single ridge in the TF plane. A more formal description of multicomponent signals is given by Cohen [48] in terms of the instantaneous frequency and instantaneous bandwidth of a signal. Consider the signal:

$$x(t) = w(t) + z(t) = A_w\mathrm{e}^{\mathrm{j}\psi_w(t)} + A_z\mathrm{e}^{\mathrm{j}\psi_z(t)}. \tag{3.15}$$

If the instantaneous bandwidths of $s(t)$ and $z(t)$ are small with respect to the difference between their instantaneous frequencies, then two "ridges" will result in the TF plane. A condition for $x(t)$ to be multicomponent can therefore be given as

$$\begin{aligned} \left|\frac{\frac{d}{dt}A_w(t)}{A_w(t)}\right| &<< \left|\tfrac{d}{dt}\psi_s(t) - \tfrac{d}{dt}\psi_z(t)\right| \\ \text{and} \quad \left|\frac{\frac{d}{dt}A_z(t)}{A_z(t)}\right| &<< \left|\tfrac{d}{dt}\psi_s(t) - \tfrac{d}{dt}\psi_z(t)\right|, \quad \forall t. \end{aligned} \tag{3.16}$$

Now let us consider the TFD of the signal $x(t)$ from Equation (3.15):

$$D_{xx}(t,\Omega) = D_{ww}(t,\Omega) + D_{wz}(t,\Omega) + D_{zw}(t,\Omega) + D_{zz}(t,\Omega). \tag{3.17}$$

The components $D_{ww}(t,\Omega)$ and $D_{zz}(t,\Omega)$ are called *auto-terms*, wile the components $D_{wz}(t,\Omega)$ and $D_{zw}(t,\Omega)$ are called *cross-terms* of the distribution of $x(t)$.

The cross-terms arise due to the bilinear nature of the TFD and can result in artificial peaks at TF locations where there is no signal energy. Often, the role of the kernel function is to reduce the effect of cross-terms. In particular, the reduced interference distributions (RID) were designed to reduce the artifacts in the TF representation while retaining good resolution of the signal auto-terms [79] for multicomponent signals.

To illuminate the discussion on multicomponent signals and TFA, a short example is given. Consider again the two-component signal $x(t)$ defined in Equation (3.15), where the phases of components $w(t)$ and $z(t)$ are given respectively by

$$\begin{aligned} \psi_w(t) &= 2\pi(0.3.t + 9.375 \times 10^{-04}.t^2 - 1.831 \times 10^{-06}.t^3) \\ \text{and} \quad \psi_z(t) &= 2\pi(0.02.t - 1.172 \times 10^{-04}.t^2 + 1.221 \times 10^{-06}.t^3). \end{aligned}$$

The TFDs of $x(t) = w(t)+z(t)$ computed using various kernel functions are shown in Figure 3.1. What can be seen from this figure is that the pseudo Wigner-Ville distribution (PWVD) contains two 'ridges' along the signal IFs, corresponding to the auto-terms of the signal components (D_{ww} and D_{zz}), as well as large peaks between these two ridges which correspond to the cross-terms (D_{wz} and D_{zw}). When comparing the PWVD to the other TFDs shown in Figure 3.1 one can see that the cross-term artifacts have been suppressed to varying degrees, at the expense of reduced auto-term resolution (broader ridges along the signal IFs).

3.1.4 Analysis of stochastic processes

The joint TF analysis of a complex stochastic process $X(t)$ may be performed using a TF spectrum, defined according to [48]:

$$D_{XX}(t,\Omega) = \frac{1}{2\pi}\int_{\mathbb{R}^2} \varphi(t-t',\tau)\mathsf{E}\left[X(t'+\tau/2)X^*(t'-\tau/2)\right]\mathrm{e}^{-\mathrm{j}\Omega\tau}dt'd\tau, \quad (3.18)$$

where it is assumed that the kernel function $\varphi(t,\tau)$ is independent of the signal. Equation (3.18) describes the average energy of a random signal at a given frequency and a given point in time. A cross-TFD between stochastic processes $X(t)$ and $Z(t)$ may also be defined analogous to Equation (3.14):

$$D_{XZ}(t,\Omega) = \frac{1}{2\pi}\int_{\mathbb{R}^2} \varphi(t-t',\tau)\mathsf{E}\left[X(t'+\tau/2)Z^*(t'-\tau/2)\right]\mathrm{e}^{-\mathrm{j}\Omega\tau}dt'd\tau, \quad (3.19)$$

We note that the cross-TFD between two zero-mean, uncorrelated random processes will always be zero.

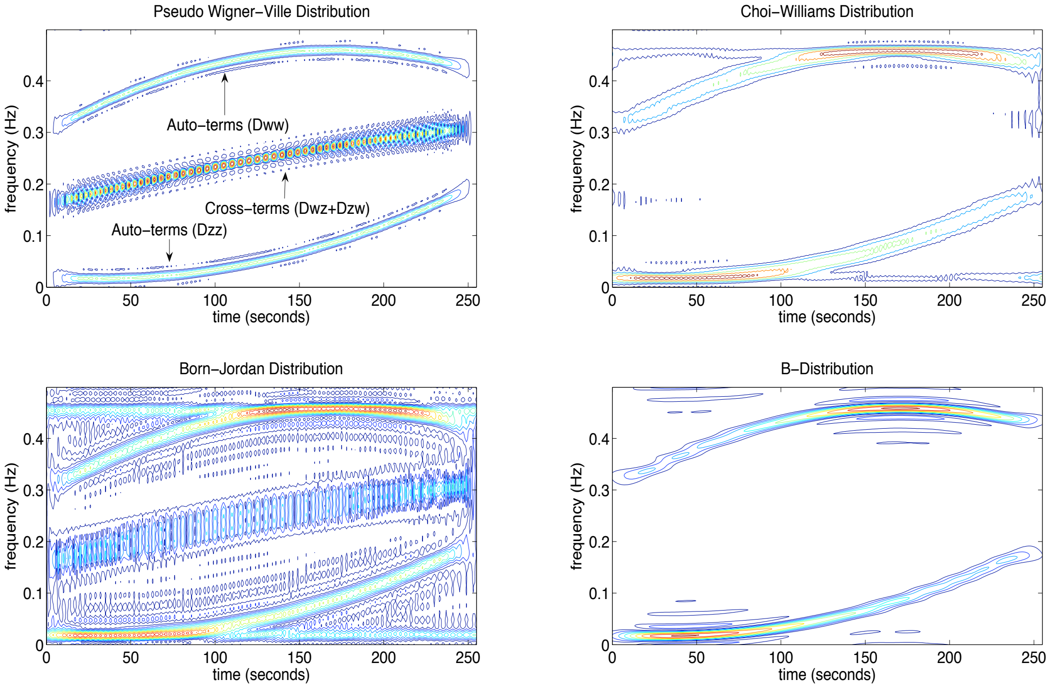


Figure 3.1: The TFDs of a multicomponent signals computed using various kernel functions.

3.2 Spatial time-frequency distributions

The analysis of single-sensor data using TFA was outlined in the previous section. The extension of this analysis to multi-variate data as in the case of sensor-array applications was proposed by Belouchrani *et al* in [19]. By forming a matrix of the auto- and cross-TFDs of the sensor signals, one can compute the so-called *spatial* TFD (STFD). Using matrix notation, the STFD of a vector signal $\boldsymbol{x}(t) = [x_1(t), \ldots, x_m(t)]^{\mathrm{T}}$ is compactly defined as

$$\boldsymbol{D}_{\boldsymbol{xx}}(t,\Omega) = \frac{1}{2\pi}\int_{\mathbb{R}^2} \boldsymbol{\Phi}(t-t',\tau)\odot\left[\boldsymbol{x}(t'+\tau/2)\boldsymbol{x}^{\mathrm{H}}(t'-\tau/2)\right]\mathrm{e}^{-\mathrm{j}\Omega\tau}dt'd\tau, \tag{3.20}$$

where the ijth element of the matrix $\boldsymbol{\Phi}(t,\tau)$ is the TF kernel function used to compute the cross-sensor TFD between sensor i and sensor j and $\odot$ denotes the Hadamard (element-wise) matrix product. To the author's knowledge, there has been no work to date looking at using different TF kernels to compute each cross-sensor TFD. In the following, a single TF kernel function, $\varphi(t,\tau)$ from Cohen's class, shall be assumed, and the STFD matrix formulated as

$$\boldsymbol{D}_{\boldsymbol{xx}}(t,\Omega) = \frac{1}{2\pi}\int_{\mathbb{R}^2} \varphi(t-t',\tau)\boldsymbol{x}(t'+\tau/2)\boldsymbol{x}^{\mathrm{H}}(t'-\tau/2)\mathrm{e}^{-\mathrm{j}\Omega\tau}dt'd\tau. \tag{3.21}$$

If the data is modelled as a vector random process, $\boldsymbol{X}(t)$, then the STFD matrix is defined by applying the expectation operator to Equation (3.21):

$$\boldsymbol{D}_{\boldsymbol{XX}}(t,\Omega) = \frac{1}{2\pi}\int_{\mathbb{R}^2} \varphi(t-t',\tau)\mathsf{E}\left[\boldsymbol{X}(t'+\tau/2)\boldsymbol{X}^{\mathrm{H}}(t'-\tau/2)\right]\mathrm{e}^{-\mathrm{j}\Omega\tau}dt'd\tau. \quad (3.22)$$

With respect to the linear model defined in Equation (1.8) for array processing, the STFD matrix may be expressed as

$$\boldsymbol{D}_{\boldsymbol{XX}}(t,\Omega) = \boldsymbol{A}\boldsymbol{D}_{\boldsymbol{ss}}(t,\Omega)\boldsymbol{A}^{\mathrm{H}} + \sigma^2\boldsymbol{I} \quad (3.23)$$

where $\boldsymbol{D}_{\boldsymbol{ss}}(t,\Omega)$ is the STFD of the source signals (assumed to be deterministic) and $\sigma^2 = \sigma_V^2(2\pi)^{-1}\int_{\mathbb{R}}\varphi(u,0)du$ is a constant determined by the noise power and the TF kernel function used. Comparing Equation (3.23) to the array covariance matrix (see Equation (2.13)), the STFD matrix retains the same 'structure'. In Equation (3.23), the auto-source TFDs (diagonal entries of $\boldsymbol{D}_{\boldsymbol{ss}}(t,\Omega)$) and the cross-source TFDs (off-diagonal entries of $\boldsymbol{D}_{\boldsymbol{ss}}(t,\Omega)$) play an analogous role to the signal auto- and cross-correlations, respectively. This important observation allows one to apply many of the conventional second-order based array processing methods to nonstationary signals by replacing the covariance matrix with the STFD matrix [7,20,22,123].

It should be noted that calculation of the STFD matrix is usually done in practice based on a discrete-time formulation[1]of Equation (3.22):

$$\boldsymbol{D}_{\boldsymbol{XX}}(n,\omega) = \sum_m\sum_l \varphi(m,l)\mathsf{E}\left[\boldsymbol{X}(n+m+l)\boldsymbol{X}^{\mathrm{H}}(n+m-l)\right]\mathrm{e}^{\mathrm{j}2\omega l}. \quad (3.24)$$

A possible estimator of Equation (3.24) given the finite-length random sample $\{\boldsymbol{X}(n)\}_{n=0}^{N-1}$ is then given by

$$\hat{\boldsymbol{D}}_{\boldsymbol{XX}}(n,\omega) = \sum_{m=-\frac{M-1}{2}}^{\frac{M-1}{2}}\sum_{l=-\frac{L-1}{2}}^{\frac{L-1}{2}} \varphi(m,l)\boldsymbol{X}(n+m+l)\boldsymbol{X}^{\mathrm{H}}(n+m-l)\mathrm{e}^{\mathrm{j}2\omega l}, \quad (3.25)$$

for $n = 0,\dots,N-1$, where M and L above are odd 'window' lengths whose maximum value is limited due to the finite observation interval. A particular estimate of $\boldsymbol{D}_{\boldsymbol{XX}}(n,\omega)$ is obtained by substituting the observations $\{\boldsymbol{x}(n)\}$ for $\{\boldsymbol{X}(n)\}$ into Equation (3.25) and is therefore denoted by $\hat{\boldsymbol{D}}_{\boldsymbol{xx}}(n,\omega)$.

3.2.1 Structure of the source STFD

The relationship given in Equation (3.23) holds for every point, (t,Ω), on the TF plane. However, the underlying structure of the matrix $\boldsymbol{D}_{\boldsymbol{ss}}$ plays a great

---

[1] A practical guide to the implementation of discrete-time TFDs can be found in [25].

role in the applicability of the STFD matrix to array processing problems. In particular, certain points on the TF plane lead to a strongly diagonal or off-diagonal structure in $\boldsymbol{D_{ss}}$ which may be exploited for direction-finding or blind source separation, as will be discussed in the following sections.

To illustrate how a particular structure in $\boldsymbol{D_{ss}}$ arises, we take the example of two FM signals impinging on a sensor array from different directions. The signals are given by:

$$\begin{aligned} s_1(t) &= \exp\{\mathrm{j}\psi(t)\} && (3.26) \\ s_2(t) &= \exp\{\mathrm{j}(\psi(t)+\omega_0 t)\}, && (3.27) \end{aligned}$$

for $t \in \mathbb{R}$. Assume that the sensor array output is sampled at $1/T_s$ samples per second yielding samples $\{x(0), \ldots, x(N-1)\}$. The STFD matrix is computed based on the discrete-time pseudo Wigner-Ville distribution (PWVD), calculated according to:

$$\hat{\boldsymbol{D}}_{\boldsymbol{xx}}(n,\omega) = \sum_{l=-(L-1)/2}^{(L-1)/2} \boldsymbol{x}(n+l)\boldsymbol{x}^{\mathrm{H}}(n-l)\mathrm{e}^{\mathrm{j}2\omega l}, \tag{3.28}$$

where n refers to the time index corresponding to time nT_s and ω is the normalized frequency in radians per second, i.e. $\omega = \Omega T_s$. The *source* STFD matrix, $\boldsymbol{D_{ss}}(n,\omega)$, is obtained by substituting $\boldsymbol{s}(n)$ for $\boldsymbol{x}(n)$ in Equation (3.28). We assume that the signal IF, $\tilde{\omega}(n)$, is approximately linear within each time interval of length LT_s and that $L << N$. For $n \in \{(L-1)/2, \ldots, N-(L-1)/2-1\}$, we obtain the following PWVDs:

$$\begin{aligned} D_{s_1s_1}(n,\omega) &\approx \frac{\sin[(\tilde{\omega}(n)-\omega)L]}{\sin(\tilde{\omega}(n)-\omega)} && (3.29) \\ D_{s_1s_2}(n,\omega) &\approx \frac{\sin[(\tilde{\omega}(n)+\omega_0/2-\omega)L]}{\sin(\tilde{\omega}(n)+\omega_0/2-\omega)}\mathrm{e}^{-\mathrm{j}\omega_0 n} && (3.30) \\ D_{s_1s_2}(n,\omega) &= D^*_{s_2s_1}(n,\omega) && (3.31) \\ D_{s_2s_2}(n,\omega) &= D_{s_1s_1}(n,\omega-\omega_0), && (3.32) \end{aligned}$$

where the above expressions become exact if $\psi(t)$ is a quadratic function of time (i.e. the signals are linear FM). The TFD at the reference sensor is plotted in Figure 3.2, where a mixture of all source auto- and cross-distributions is observed. From Figure 3.2 one can clearly see the constant peaks due of the auto-source and oscillatory peaks of the cross-source distributions.

Let us now examine the particular regions of interest on the TF plane, namely the locations of auto- and cross-source peaks. From expressions (3.29) and (3.32)

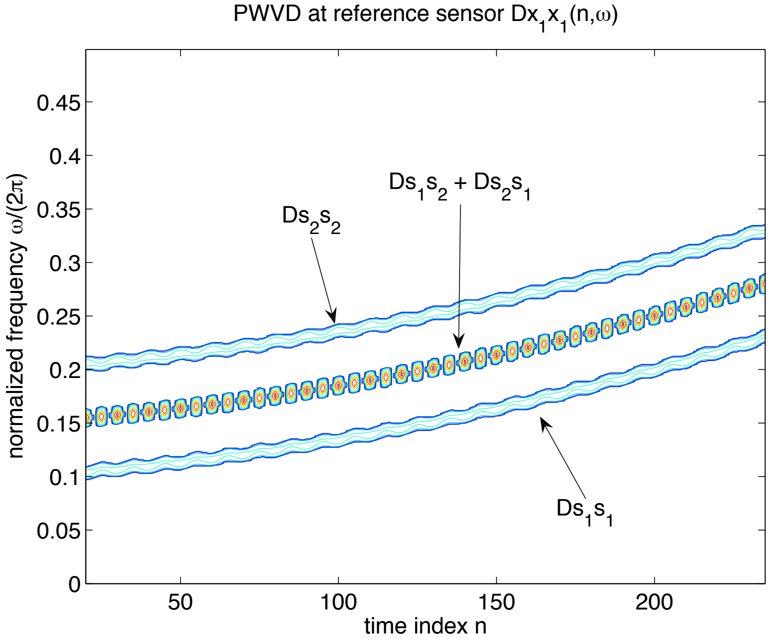


Figure 3.2: PWVD of a mixture of two FM signals as observed at a sensor array.

we see that the auto-source peaks of sources one and two occur at TF locations $(n, \tilde{\omega}(n))$ and $(n, \tilde{\omega}(n) + \omega_0)$ respectively, and are of constant amplitude L. On the other hand, the cross-source peaks are oscillatory but exhibit the largest magnitude along the TF trajectory $(n, \tilde{\omega}(n) + \omega_0/2)$. In these cases, the source STFD matrix takes the form:

$$\boldsymbol{D_{ss}}(n, \tilde{\omega}(n)) = \begin{bmatrix} L & \epsilon_{12} \\ \epsilon_{21} & \epsilon_{22} \end{bmatrix} \tag{3.33}$$

$$\boldsymbol{D_{ss}}(n, \tilde{\omega}(n) + \omega_0) = \begin{bmatrix} \epsilon_{11} & \epsilon_{12} \\ \epsilon_{21} & L \end{bmatrix} \tag{3.34}$$

$$\boldsymbol{D_{ss}}(n, \tilde{\omega}(n) + \omega_0/2) = \begin{bmatrix} \epsilon_{11} & L\mathrm{e}^{-\mathrm{j}\omega_0 n} \\ L\mathrm{e}^{\mathrm{j}\omega_0 n} & \epsilon_{22} \end{bmatrix} \tag{3.35}$$

where ϵ_{ij} denotes a small value with magnitude much less than L. What is clear from (3.33)-(3.35) is that the matrix $\boldsymbol{D_{ss}}$ exhibits diagonal and off-diagonal structure at auto- and cross-source peaks respectively[2]. It is this particular structure which gives rise to powerful array processing methods for solving BSS and DOA estimation problems. The application of STFD matrices to these problems is

---

[2] In the example given here, we have considered a simple case with parallel TF signatures which do not directly overlap. If the TF signatures do cross, then the cross-source peaks occur at the same TF locations as the auto-source peaks, and there is no guarantee of a particular (off-)diagonal structure in $\boldsymbol{D_{ss}}$ at these points.

reviewed in the following sections.

3.2.2 Blind source separation using STFD matrices

A number of BSS algorithms have been proposed, which make use of the STFD matrices discussed in the previous section. [18, 54, 66]. The various schemes all exploit the underlying diagonal or off-diagonal structure of STFD matrices at certain locations on the TF plane. BSS is achieved by first constructing a set of STFD matrices, followed by joint diagonalization (JD), joint off-diagonalization (JOD) or combined JD/OD, to estimate the mixing system. The optimization of JD/OD criteria has been proposed using both orthogonal [18] and non-orthogonal [54, 121] constraints.

In the following, a brief review of different STFD-based approaches to BSS are given. For notational convenience, we shall denote a point, (t, Ω), on the TF plane by the symbol $\boldsymbol{\zeta}$. Further, it is assumed that there are two known sets of TF points, $\mathcal{S}_D$ and $\mathcal{S}_O$, at which the matrix $\boldsymbol{D_{ss}}(\boldsymbol{\zeta})$ has a strong diagonal or off-diagonal structure respectively. Based on Equation (3.23), one can define a “noise compensated” STFD matrix as:

$$\underline{\boldsymbol{D}}_{\boldsymbol{XX}}(\boldsymbol{\zeta}) \triangleq \boldsymbol{D}_{\boldsymbol{XX}}(\boldsymbol{\zeta}) - \sigma^2 \boldsymbol{I} = \boldsymbol{A}\boldsymbol{D_{ss}}(\boldsymbol{\zeta})\boldsymbol{A}^{\mathrm{H}}. \tag{3.36}$$

For $\boldsymbol{\zeta} \in \mathcal{S}_D$ or $\boldsymbol{\zeta} \in \mathcal{S}_O$, the matrix $\boldsymbol{A}^{-1}$ is a diagonalization or off-diagonalizer of $\underline{\boldsymbol{D}}_{\boldsymbol{XX}}(\boldsymbol{\zeta})$, respectively. As defined in Section 2.1.1, the goal of BSS is to find an unmixing matrix, $\boldsymbol{B}$, which is similar to $\boldsymbol{A}^{-1}$ (up to scaling and permutation indeterminacy).

Orthogonal JD/OD

Orthogonal JD of STFD matrices was first proposed in [20], and later extended to orthogonal JD/OD in [18]. This approach requires the mixing matrix to be orthogonal, which in general is not the case. A whitening step is therefore first applied to the data, to fulfill the orthogonality constraint. A matrix $\boldsymbol{W}$ is found, such that $\boldsymbol{W}\boldsymbol{A} = \boldsymbol{U}$ where $\boldsymbol{U}$ is orthogonal. One then forms two sets of matrices:

$$\begin{aligned} \underline{\mathcal{M}} &= \{\boldsymbol{M}_k = \boldsymbol{W}\underline{\boldsymbol{D}}_{\boldsymbol{XX}}(\boldsymbol{\zeta}_k)\boldsymbol{W}^{\mathrm{H}};\ \boldsymbol{\zeta}_k \in \mathcal{S}_D\}_{k=1}^{M} \\ \underline{\mathcal{N}} &= \{\boldsymbol{N}_k = \boldsymbol{W}\underline{\boldsymbol{D}}_{\boldsymbol{XX}}(\boldsymbol{\zeta}_k)\boldsymbol{W}^{\mathrm{H}};\ \boldsymbol{\zeta}_k \in \mathcal{S}_O\}_{k=1}^{N} \end{aligned} \tag{3.37}$$

The matrices in $\underline{\mathcal{M}}$ and $\underline{\mathcal{N}}$ are jointly diagonalized and off- diagonalized by the same, unknown orthogonal mixing matrix $\boldsymbol{U}$. To perform orthogonal JD/OD the following criterion was proposed in [18], which can be optimized using an efficient iterative Jacobi-like technique:

$$C_1(\boldsymbol{V};\underline{\mathcal{M}},\underline{\mathcal{N}}) = \sum_{i=1}^{m}\left(\sum_{k=1}^{M}|\boldsymbol{v}_i^H\boldsymbol{M}_k\boldsymbol{v}_i|^2 - \sum_{k=1}^{N}|\boldsymbol{v}_i^H\boldsymbol{N}_k\boldsymbol{v}_i|^2\right)$$

where $\boldsymbol{V} = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n]$ is an orthogonal matrix. BSS based on C_1 is summarized in Table 3.3

1. Obtain an estimate, $\hat{\boldsymbol{W}}$, of the whitening transform[a].
2. Form the sets of STFD matrices $\underline{\mathcal{M}}$ and $\underline{\mathcal{N}}$.
3. Estimate the orthogonal mixing matrix: $\hat{\boldsymbol{U}} = \arg\max_{\boldsymbol{V}} C_1(\boldsymbol{V};\underline{\mathcal{M}},\underline{\mathcal{N}})$
4. Estimate the separating matrix: $\hat{\boldsymbol{B}} = \hat{\boldsymbol{U}}^{\mathrm{H}}\hat{\boldsymbol{W}}$

---

[a] Estimation of $\boldsymbol{W}$ from the observations is discussed in [20] and reviewed in Section 2.1.3 of this dissertation.

Table 3.3: Orthogonal JD/OD algorithm for BSS using STFD matrices.

Non-Orthogonal JD/OD

As noted in [33], the whitening transform used in orthogonal methods limits the achieveable BSS performance. It is therefore of interest to use non-orthogonal approaches. In this case, two sets of matrices are formed from the unwhitened STFDs:

$$\begin{aligned}\mathcal{M} &= \{\boldsymbol{M}_k = \boldsymbol{D}_{\boldsymbol{XX}}(\boldsymbol{\zeta}_k);\ \boldsymbol{\zeta}_k \in \mathcal{S}_D\}_{k=1}^{M} \\ \mathcal{N} &= \{\boldsymbol{N}_k = \boldsymbol{D}_{\boldsymbol{XX}}(\boldsymbol{\zeta}_k);\ \boldsymbol{\zeta}_k \in \mathcal{S}_O\}_{k=1}^{N}\end{aligned} \tag{3.38}$$

which are jointly diagonalized and off-diagonalized respectively by the same, unknown mixing matrix $\boldsymbol{A}$. Two non-orthogonal JD/OD methods have been proposed in the literature.

The first approach is from [121], which implements JD only, and therefore makes use only of $\mathcal{M}$. The JD criterion takes the form

$$C_2(\boldsymbol{C},\mathcal{L};\mathcal{M}) = \sum_{k=1}^{M}\gamma_k\|\boldsymbol{M}_k - \boldsymbol{C}\boldsymbol{\Lambda}_k\boldsymbol{C}^H\|_F^2 \tag{3.39}$$

where $\{\gamma_k\}_{k=1}^{M}$ are positive weights and $\mathcal{L} = \{\boldsymbol{\Lambda}_k\}_{k=1}^{M}$ are diagonal matrices. The matrix $\boldsymbol{C}$ corresponds to an estimate of the mixing matrix $\boldsymbol{A}$. Iterative optimization of Equation (3.39) was proposed in two steps: In the first step C_2 is minimized w.r.t. to the l-th column of $\boldsymbol{C}$, in the second step C_2 is minimized w.r.t. $\boldsymbol{\mathcal{L}}$. It should be noted that this approach only holds when there are the same number of sensors as sources ($m = n$). A BSS algorithm based on C_2 was proposed in [66] and is summarized in Table 3.4.

1. Form the set of STFD matrices $\mathcal{M}$.
2. Estimate the non-orthogonal mixing matrix:
   $\hat{\boldsymbol{A}} = \arg\min_{\boldsymbol{C}} C_2(\boldsymbol{C}, \mathcal{L}; \mathcal{M})$
3. Estimate the separating matrix: $\hat{\boldsymbol{B}} = \hat{\boldsymbol{A}}^{-1}$

Table 3.4: Non-orthogonal JD algorithm for BSS using STFD matrices.

A second approach to non-orthogonal JD/OD was proposed in [54] via minimization of the following criterion:

$$C_3(\boldsymbol{B}; \mathcal{M}, \mathcal{N}) = \sum_{l}^{m} \boldsymbol{b}_l \boldsymbol{R}_l(\boldsymbol{B}) \boldsymbol{b}_l^H \tag{3.40}$$

where

$$\begin{aligned}
\boldsymbol{R}_l(\boldsymbol{B}) &= \alpha \boldsymbol{R}_{D,l}(\boldsymbol{B}) + (1-\alpha)\boldsymbol{R}_{Z,l}(\boldsymbol{B}) \\
\boldsymbol{R}_{D,l}(\boldsymbol{B}) &= \sum_{i=1}^{M} \boldsymbol{M}_i \sum_{k=1,k\neq l}^{m} (\boldsymbol{b}_k^H \boldsymbol{b}_k) \boldsymbol{M}_i \\
\boldsymbol{R}_{Z,l}(\boldsymbol{B}) &= \sum_{i=1}^{N} \boldsymbol{N}_i \boldsymbol{b}_l^H \boldsymbol{b}_l \boldsymbol{N}_i
\end{aligned}$$

with $0 \leq \alpha \leq 1$ and $\boldsymbol{B} = [\boldsymbol{b}_1, \ldots, \boldsymbol{b}_m]^T$. Minimization of Equation (3.40) was proposed in an iterative row-wise scheme. The BSS algorithm based on C_3 is summarized in Table 3.5.

1. Form the sets of matrices $\mathcal{M}$ and $\mathcal{N}$.
2. Estimate the separating matrix:
   $\hat{\boldsymbol{B}} = \arg\min_{\boldsymbol{B}} C_3(\boldsymbol{B}; \mathcal{M}, \mathcal{N})$

Table 3.5: Non-orthogonal JD/OD algorithm for BSS using STFD matrices.

Orthogonal vs. non-orthogonal schemes

While the orthogonal JD/OD schemes provide a potential performance improvement over the non-orthogonal schemes, this point alone does not give the overall picture of which scheme would be 'best' in a given application. In Appendix 3.A, a more detailed comparison of the orthogonal and non-orthogonal schemes is conducted in terms of BSS performance, computational complexity and robustness to the correct choice of TF points.

In general it was found that, for high SNR, the non-orthogonal algorithms could achieve superior performance to the orthogonal approach, at the cost of higher computational complexity. The non-orthogonal JD-only algorithm summarized in Table 3.4 was the most computationally expensive of the schemes, though it proved slightly more robust than the other non-orthogonal scheme, to erroneous TF point selection. The orthogonal JD/OD algorithm was found to give superior performance at low to mid SNR, at the lowest computational complexity, while being the most robust method overall to the inclusion of falsely chosen TF points. Given the above considerations, the orthogonal JD/OD approach is very attractive, especially for noisy environments, and shall be re-visited in the following chapter.

Choice of the TF kernel

The STFD matrices required for BSS, as defined in Equations (3.37) and (3.38) may be calculated according to Equation (3.21), using an arbitrary TF kernel function. There are, however, certain disadvantages in using the TF smoothing kernels discussed in Section 3.1.3, with respect to the BSS problem. The use of the smoothing kernels generally reduces the resolution of the auto-terms, while suppressing the cross-terms peaks. This effect is not desirable for either JD or JOD as it degrades the desired matrix structure which is exploited for BSS. Some general comments on the use of TF smoothing kernels for BSS have also been made in [122]. The main points raised are as follows:

1. The effectiveness of smoothing kernels at reducing cross-terms is highly impaired in the case of signals with closely spaced TF signatures.

2. Smoothing kernels are not effective at reducing cross-terms in all regions of the TF plane.

3. Some kernels displace cross-terms to reside over auto-term locations, violating the diagonal nature of STFDs which is required for source separation.

4. When finite observation periods are considered, leakage effects will generally cause main and/or side-lobes of cross-terms to intrude on auto-term regions and this cannot be prevented or entirely removed.

Given the above reasoning, we focus on the pseudo Wigner-Ville (PWVD) distribution in the following, which retains good resolution of both the auto- and cross-term peaks, provided the IF is approximately linear within all time sub-intervals equal to the PWVD window length L.

3.2.3 Direction-of-arrival estimation

Solving the direction finding problem using STFD matrices has been considered in [7, 22, 123, 124]. As illustrated by Equation (3.23), the STFD matrix exhibits the same Eigen-structure as the sensor covariance matrix. This fact has facilitated the direct application of many conventional direction finding methods to STFD matrices. In the following, a brief review of two techniques for direction-finding based on STFD matrices is provided. The first approach is based on the subspace decomposition of an averaged STFD matrix and the second on joint block-diagonalization of a number of STFD matrices.

For clarity of the presentation, the data is modelled defined in Section 2.2, is repeated here:

$$\boldsymbol{X}(t) = \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{s}(t) + \boldsymbol{V}(t), \tag{3.41}$$

where the source signal vector is given by Equation (3.42). Given observations $\{\boldsymbol{x}(n)\}_{n=1}^{N}$, of $\boldsymbol{X}(t)$, the goal is to estimate $\boldsymbol{\theta}$.

Subspace estimation via averaged STFD matrices

As discussed in Section 2.2, the popular MUSIC estimator performs direction finding based on the decomposition of the sensor covariance matrix into signal and noise subspaces. Subspace analysis of the STFD matrix can therefore provide useful insight into the performance characteristics of STFD-based estimators. Constant amplitude FM source signals were considered in [124] due to the clear TF signature they exhibit:

$$\boldsymbol{s}(t) = [s_1(t), \ldots, s_d(t)]^{\mathrm{T}} = [A_1 \mathrm{e}^{\mathrm{j}\psi_1(t)}, \ldots, A_d \mathrm{e}^{\mathrm{j}\psi_d(t)}]^{\mathrm{T}} \tag{3.42}$$

where source k has constant amplitude A_k and time-varying phase $\psi_k(t)$.

The analysis done in [124] was based on using the discrete-time PWVD to calculate the STFD matrices, as defined in Equation (3.28). The results assume that the source signals are mutually uncorrelated over the observation period and have negligible third-order derivative of the phase over the PWVD window length L. Under these assumptions and from Equation (3.28), the auto-source TFDs along the signal IFs are approximately

$$D_{s_k s_k}(n, \tilde{\omega}_k(n)) \approx L|A_k|^2, \quad n \in \mathcal{S}_{N_o}, \tag{3.43}$$

where $\mathcal{S}_{N_o} = \{(L-1)/2, \dots, N-(L-1)/2-1\}$. By substituting Equation (3.43) into Equation (3.23) the array output STFD is given by

$$\mathsf{E}\left[\hat{\boldsymbol{D}}_{\boldsymbol{XX}}(n, \tilde{\omega}_k(n))\right] \approx L|A_k|^2 \boldsymbol{a}(\theta_k)\boldsymbol{a}^{\mathrm{H}}(\theta_k) + \sigma^2 \boldsymbol{I} \tag{3.44}$$

for $k = 1, \dots, d$. Let us define a subset $\mathcal{S}_{d_o} \subseteq \{1, \dots, d\}$ of the source indices containing $d_o = \#\mathcal{S}_{d_o} \leq d$ elements. Averaging $\hat{\boldsymbol{D}}_{\boldsymbol{XX}}(n, \omega)$ over the instantaneous frequency of d_o signals and over the $N_o = \#\mathcal{S}_{N_o} = N - L + 1$ points of constant distribution, gives an 'averaged' STFD matrix:

$$\hat{\boldsymbol{D}} = \frac{1}{d_o N_o} \sum_{q \in \mathcal{S}_{d_o}} \sum_{n \in \mathcal{S}_{N_o}} \hat{\boldsymbol{D}}_{\boldsymbol{XX}}(n, \tilde{\omega}_q(n)). \tag{3.45}$$

Substituting the result of Equation (3.44) into Equation (3.45) gives the expectation of the averaged STFD matrix:

$$\boldsymbol{D} \triangleq \mathsf{E}\left[\hat{\boldsymbol{D}}\right] = \frac{L}{d_o} \boldsymbol{A}^o \boldsymbol{S}^o (\boldsymbol{A}^o)^{\mathrm{H}} + \sigma^2 \boldsymbol{I}, \tag{3.46}$$

where $\boldsymbol{A}^o$ is the mixing matrix formed with steering vectors $\{\boldsymbol{a}_k\}_{k \in \mathcal{S}_{d_o}}$ and $\boldsymbol{S}^o$ is a diagonal matrix with diagonal elements $\{|A_k|^2\}_{k \in \mathcal{S}_{d_o}}$.

It follows from Equation (3.46) that the averaged STFD and the array output covariance matrix have identical signal and noise subspaces, for the particular d_o signals for which the TF averaging is performed. Further, in [124] it is shown that the eigenvalues of $\boldsymbol{D}$ are given by

$$\lambda_i^{tf} = \begin{cases} \frac{L}{d_o}\lambda_i^o + \left(1 - \frac{L}{d_o}\right)\sigma^2, & i \leq d_o \\ (\sigma^{tf})^2 = \sigma^2, & d_o < i \leq m \end{cases} \tag{3.47}$$

where λ_i^o is the eigenvalue corresponding to the ith source obtained from Eigen decomposition of the array covariance matrix and λ_i^{tf} is the corresponding eigenvalue obtained using Equation (3.46). This implies that the largest d_o eigenvalues are amplified using STFD analysis. These results have a significant consequence; any subspace-based methods for array signal processing which use the spatial covariance matrix may be directly applied to the averaged STFD matrix instead, with a gain in performance. The best gain in performance is obtained by selecting $d_o = 1$ and performing the TF averaging over each individual source for direction-finding. Further, by selecting $d_o < d$ sources, the number of sensors required for indentifiability of the DOA estimation problem is reduced.

To illustrate the effectiveness of using the averaged STFD matrix for direction finding, a short example is given here. We consider two linear FM signals impinging on a uniform linear array of 6 sensors having DOAs of $-2°$ and $3°$ w.r.t. the array broadside respectively. The MUSIC spectrum is calculated according to Equation (2.40), where the noise subspace is estimated using both the spatial covariance matrix and the averaged STFD matrix. TF averaging over all sources ($d_o = d = 2$) and over each source individually ($d_o = 1$) is considered. Several realizations of the MUSIC spectrum at an SNR of -15 dB are shown in Figure 3.3. It can be seen that, while covariance-based MUSIC fails to properly resolve both sources, TF-MUSIC is still successful, with the best resolution obtained for $d_o = 1$.

Subspace estimation via joint block-diagonalization

The concept of joint block diagonalization (JBD) for estimating signal and noise subspaces was introduced by Belouchrani *et al* with application to direction finding in correlated noise fields [21]. This method was later applied to STFD matrices for DOA estimation [22]. A singular value decomposition (SVD) of the mixing matrix yields

$$\boldsymbol{A} = [\boldsymbol{U_s} \;\; \boldsymbol{U_n}][\boldsymbol{S} \;\; \boldsymbol{0}]\boldsymbol{V}^{\mathrm{H}} \tag{3.48}$$

which can be combined with Equation (3.23) to show that:

$$\underline{\boldsymbol{D}}_{\boldsymbol{XX}}(\boldsymbol{\zeta}) = [\boldsymbol{U_s} \;\; \boldsymbol{U_n}]\boldsymbol{S}(\boldsymbol{\zeta})[\boldsymbol{U_s} \;\; \boldsymbol{U_n}]^{\mathrm{H}}. \tag{3.49}$$

In the above, $\boldsymbol{S}(\boldsymbol{\zeta})$ is a block diagonal matrix given by

$$\boldsymbol{S}(\boldsymbol{\zeta}) = [\boldsymbol{S}\boldsymbol{V}^{\mathrm{H}}\boldsymbol{D_{ss}}(\boldsymbol{\zeta})\boldsymbol{V}\boldsymbol{S} \;\; \boldsymbol{0}].$$

The matrices $\boldsymbol{E_s}$ and $\boldsymbol{E_n}$ are independent of the TF location and span the signal and noise subspaces respectively, allowing block diagonalization of any matrix $\underline{\boldsymbol{D}}_{\boldsymbol{XX}}(\boldsymbol{\zeta})$ by the unitary transform $\boldsymbol{E} = [\boldsymbol{E_s} \;\; \boldsymbol{E_n}]$.

Rather than simply estimating $\boldsymbol{E_s}$ and $\boldsymbol{E_n}$ from a single matrix $\hat{\underline{\boldsymbol{D}}}_{\boldsymbol{xx}}(\boldsymbol{\zeta})$, it is proposed in [22] to perform JBD of a set $\{\hat{\underline{\boldsymbol{D}}}_{\boldsymbol{xx}}(\boldsymbol{\zeta}_k)\}_{k=1}^{K}$ of STFD matrices. This avoids indeterminacy problems which can arise when $\boldsymbol{D_{ss}}(\boldsymbol{\zeta})$ is singular. JBD is

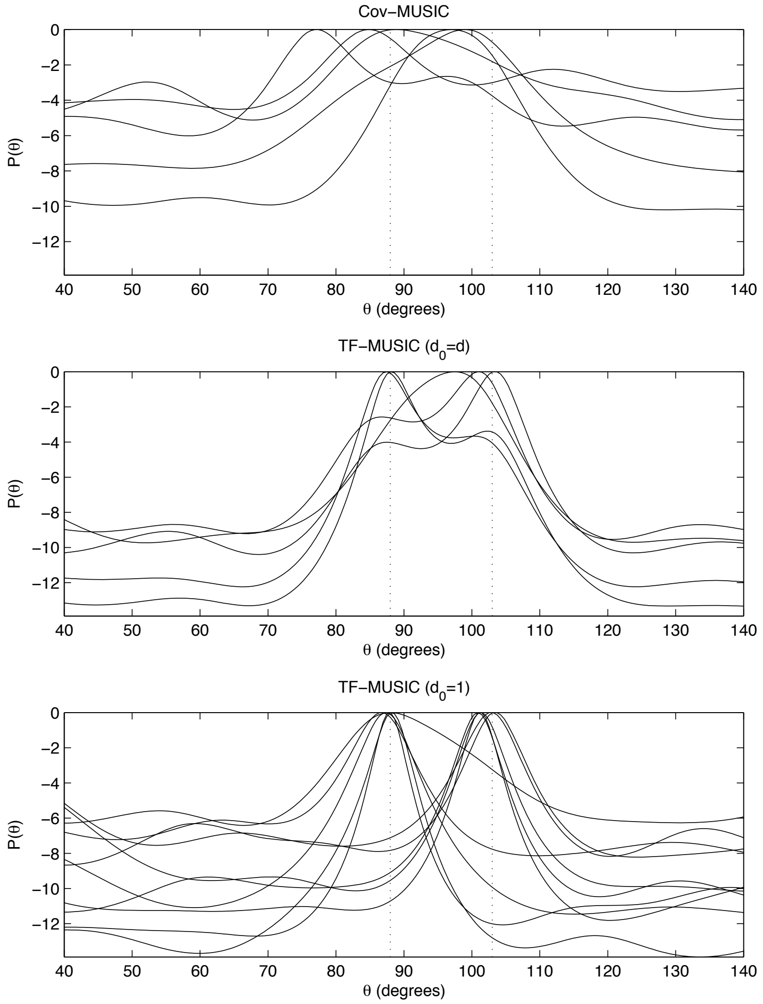


Figure 3.3: Several realisations of the MUSIC spectrum computed using the array covariance matrix (top) and the averaged STFD matrix for $d_o = 2$ (middle) and $d_o = 1$ (bottom).

achieved by maximization under unitary transform of the criterion

$$C(\boldsymbol{U}) = \sum_{k=1}^{K} \sum_{i,j=1}^{m} |\boldsymbol{u_i}^{\mathrm{H}} \hat{\underline{\boldsymbol{D}}}_{\boldsymbol{xx}}(\boldsymbol{\zeta}_k) \boldsymbol{u_j}|^2 \tag{3.50}$$

over the set of unitary matrices $\boldsymbol{U} = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m]$. Use of JBD to estimate the signal and noise subspace was shown in [7] to be more robust to inclusion of cross-term locations than the averaged STFD.

3.3 Summary

This chapter has provided an introduction to the topic of time-frequency analysis and a brief review of how, through the use of spatial time-frequency distributions, this powerful signal analysis tool may be applied to array processing problems.

With respect to the blind source separation problem, the STFD solution involves joint diagonalization or off-diagonalization of a number of STFD matrices. These matrices must be constructed at TF locations corresponding to source auto- and cross-term peaks respectively. While the JD/OD may be done with or without an orthogonality constraint, the orthogonal approach was found to have lower computational complexity, better performance at low SNR and be more robust against incorrect selection of STFD matrices. A critical step in the BSS algorithm, not discussed in detail in this chapter, is the procedure for automatically selecting the TF locations at which the STFD matrices for JD/OD are constructed. This problem shall be addressed in the following chapter of this dissertation.

When solving the direction finding problem, the use of averaged STFD matrices was shown via simulations to provide improved performance over covariance-based techniques, especially at low SNR and for closely spaced signals. By averaging over TF signatures of a subset of the source signals, the dimensionality of the direction finding problem may also be reduced. The best performance could be achieved by averaging over the TF signature of each source individually. If the TF signatures of the signal are not know *a priori*, then they must first be estimated before one can construct the averaged STFD matrix for direction finding. The problem of TF signature estimation is addressed in Chapter 5 of this dissertation.

3.A Comparison of orthogonal and non-orthogonal JD/OD for BSS

This appendix provides a comparison of the BSS methods discussed in Section 3.2.2 in terms of BSS performance, computational complexity and robustness to incorrect TF point selection. The algorithms summarized in Tables 3.3, 3.4 and 3.5 shall be referred to as *Bel.*, *Yer.* and *Fad.* respectively.

The comparison is aided by simulation examples, for which the following source signals defined:

$$\begin{aligned} s_1(t) &= \beta_1 e^{j\cdot 2\pi \cdot (0.05\cdot t + 0.625\cdot 10^{-3}\cdot t^2)} \\ s_2(t) &= \beta_2 e^{j\cdot 2\pi \cdot (0.3\cdot t + 0.58594\cdot 10^{-3}\cdot t^2)} \\ s_3(t) &= \beta_3 e^{j\cdot 2\pi \cdot (0.11\cdot t + 2.1484\cdot 10^{-3}\cdot t^2 - 6.9\cdot 10^{-6}\cdot t^3)}. \end{aligned}$$

Note that the source signals include both overlapping and non-overlapping TF components. In order to show results comparing all algorithms, we used an example where $m = n = 3$. The mixing matrix was:

$$\boldsymbol{A} = \begin{pmatrix} 1 & 0.5 & 0.1 \\ -0.5 & 1 & -0.75 \\ 0 & 0.2 & 1 \end{pmatrix} \tag{3.51}$$

The observations were obtained by taking 256 samples of $\boldsymbol{X}(t)$, with $\boldsymbol{V}(t)$ generated as a complex circular white Gaussian noise process. The SNR was defined as $= \beta^2/\sigma_v^2$, where $\beta = \beta_1 = \beta_2 = \beta_3$. The STFD matrix was computed using a discrete pseudo Wigner-Ville distribution (PWVD) with a Hamming window of length 33. The algorithms were evaluated using the following BSS performance index [54]:

$$\begin{aligned} I(\hat{\boldsymbol{B}}\boldsymbol{A}) = \frac{1}{N(N-1)} \Bigg[&\sum_{i=1}^{N} \left(\sum_{j=1}^{N} \frac{|(\hat{\boldsymbol{B}}\boldsymbol{A})_{ij}|^2}{\max_l |(\hat{\boldsymbol{B}}\boldsymbol{A})_{il}|^2} \right) - 1 \\ &+ \sum_{j=1}^{N} \left(\sum_{i=1}^{N} \frac{|(\hat{\boldsymbol{B}}\boldsymbol{A})_{ij}|^2}{\max_l |(\hat{\boldsymbol{B}}\boldsymbol{A})_{lj}|^2} \right) - 1 \Bigg], \end{aligned} \tag{3.52}$$

which measures the average power of interfering sources present in each separated source.

3.A.1 Performance vs SNR

The BSS performance of the algorithms was simulated with respect to SNR, with $M = N = 60$. The results shown in Figure 3.4 indicate that *Bel.* is better for low to mid SNR. This result is significant because in the previous presentation of non-orthogonal schemes [54, 55, 66], no noise was considered. Clearly, for very high SNR, the performance of the orthogonal scheme becomes limited by the whitening bound, however in many practical situations where the noise is significant, the orthogonal JD/OD approach can, in fact, demonstrate superior BSS performance.

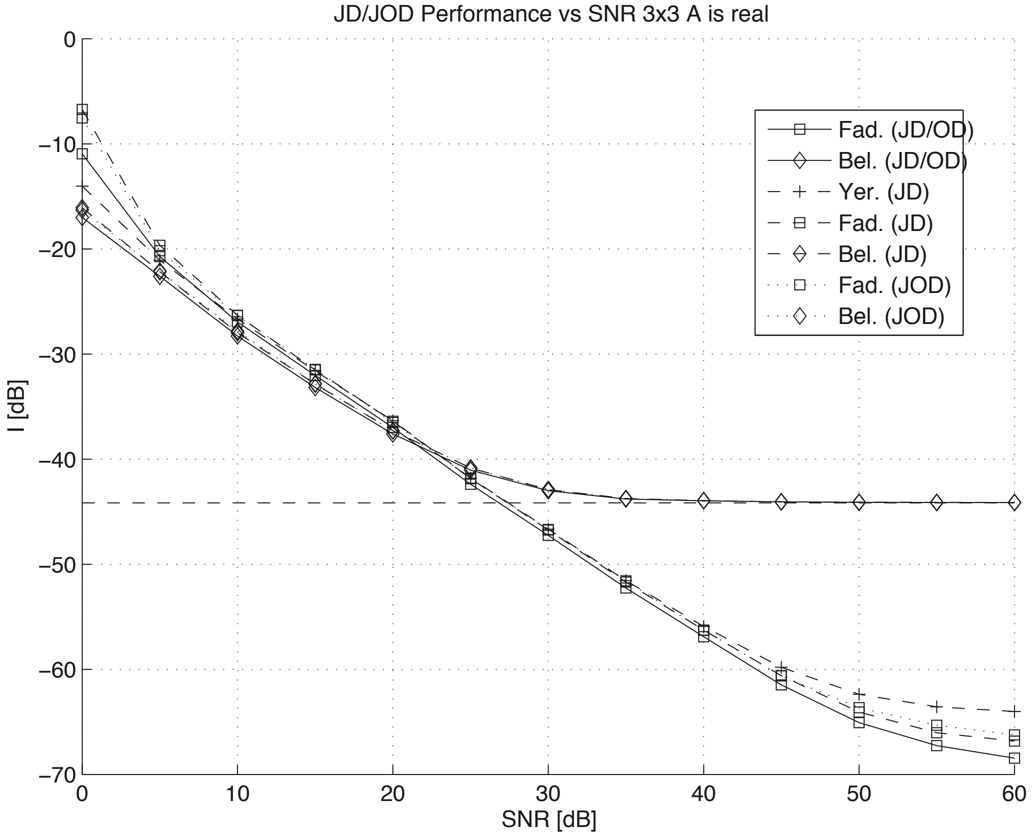


Figure 3.4: BSS performance vs SNR.

3.A.2 Computational Complexity

With respect to computational complexity, there are three main factors: The cost of computing the set of STFD matrices, the cost of computing the JD/OD criterion for each iteration and the number of iterations required for convergence. The cost of computing TFDs depends on the choice of distribution and the number of observations, which is assumed common to all schemes, and not discussed further here[3]. For the orthogonal scheme, there is also the initial cost of computing the whitening transform. The cost per iteration to compute the JD/OD

---

[3] Complexity and computationally efficient implementations of TFDs are discussed in [5, 50, 51, 113].

criteria, in terms of the order of the number of floating point operations required, is summarized in Table 3.6.

Algorithm	Whitening	Cost per Iteration
Bel.	$O(m^3)$	$O((M+N)\cdot n^3)$
Fad.	–	$O((M+N)\cdot m^3)$
Yer.	–	$O(M\cdot m^4)$

Table 3.6: Comparison of the computational complexity of the BSS algorithms.

It is important to note that, while the orthogonal scheme has a once-off cost for computing the whitening transform, this is insignificant with respect to the cost per-iteration of optimizing the JD/OD criterion. Further, in cases where there are more sensors than sources, the whitening transform reduces the dimension of the optimization problem from $O(m^3)$ to $O(n^3)$, implying a significant computational saving compared with the non-orthogonal approach, as the number of sensors increases.

It was found that all algorithms achieve convergence at about the same number of iterations, when $m = n$. It was also found that the convergence was not significantly effected by the SNR. These points are illustrated by an example where the BSS performance index versus the number of iterations was simulated for SNRs of 5, 20 and 40 dB. The results are shown in Figure 3.5.

3.A.3 Robustness to TF Point Selection

The robustness of each algorithm with respect to the choice of TF points was also considered. For this study, we have simulated the BSS performance of each algorithm versus the number of points used from $\mathcal{S}_D$ or $\mathcal{S}_O$, when a number of falsely selected points are included. The false points are chosen from the TF locations where there are strong peaks simultaneously in *both* auto- and cross-source distributions, which are likely to be falsely selected. At these locations, the source STFD does not have a particular diagonal or off-diagonal structure.

The simulations were conducted for an SNR of 20 dB, choosing 0, 1 and 3 false points. The results of JD and JD/OD are shown in Figure 3.6. It can be seen that the *Bel.* algorithm (orthogonal JD/OD) is the most robust to inclusion of false TF points. Combined JD/OD approaches were, in general, more robust to false PS, while JOD-only approaches (not shown) were the least robust. It is interesting to note that for the orthogonal *Bel.* algorithm, using a combined

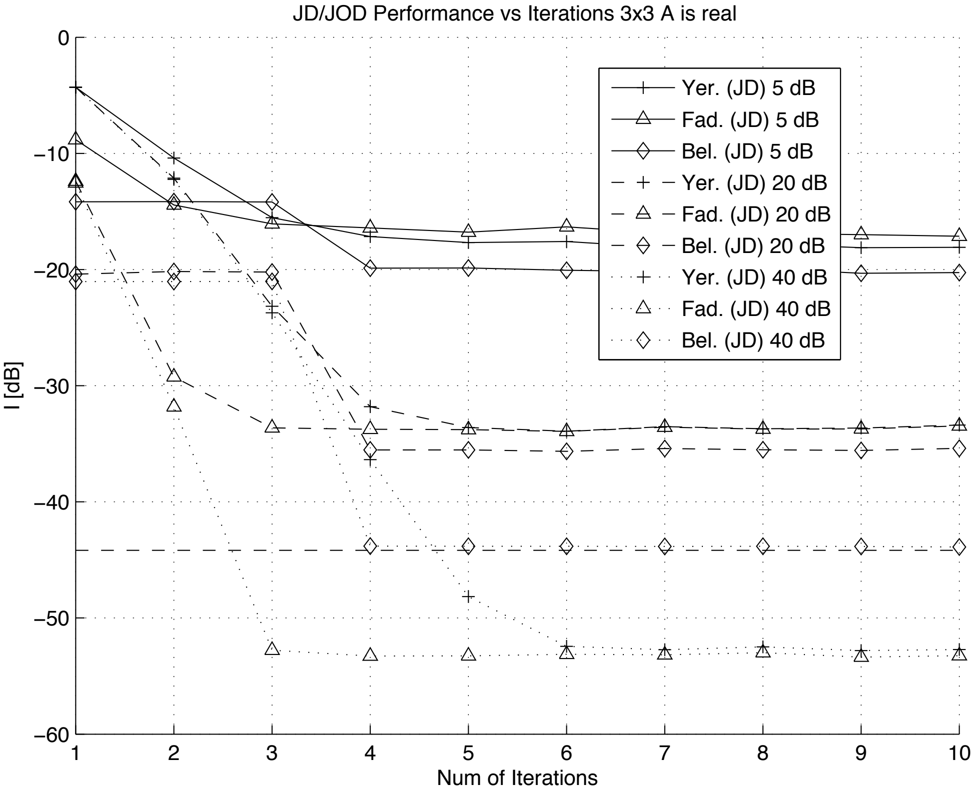
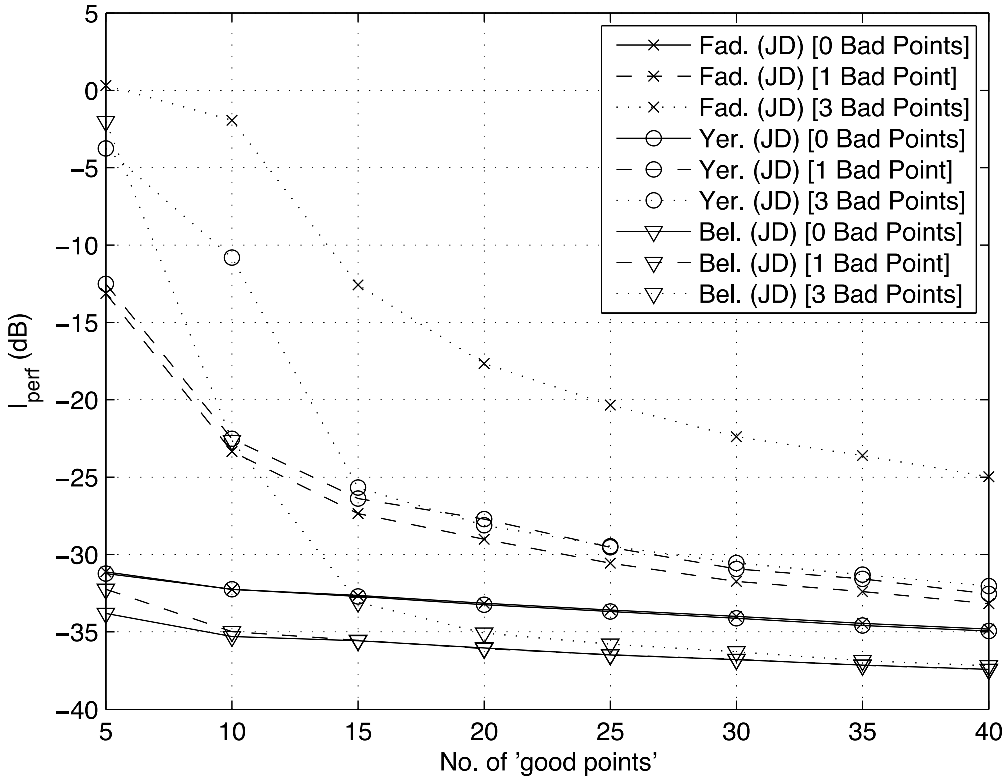


Figure 3.5: BSS performance vs iterations.

JD/OD approach meant that the inclusion of false TF points did not degrade BSS performance significantly.



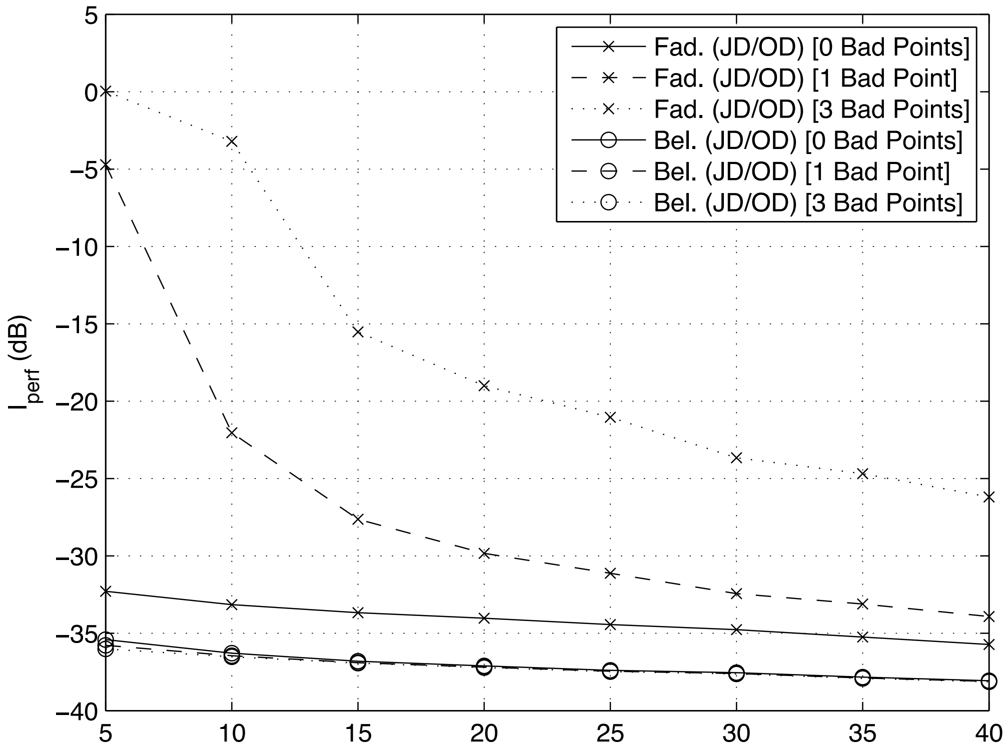


Figure 3.6: BSS performance vs number of TF points for JD only (top) and combined JD/OD (bottom).

Chapter 4

Time-Frequency Point Selection

Based on the review of time-frequency array processing methods given in the previous chapter, the problem of TF point selection, i.e. determining which locations on the TF plane correspond to auto- or cross-source peaks, is of great importance. One major component of this dissertation deals with the point selection problem.

4.1 Introduction

The use of STFD matrices for BSS was reviewed in the previous chapter and shown to provide excellent performance for separation of nonstationary sources. This approach requires joint diagonalization (JD) or off-diagonalization (JOD) of a set of STFD matrices formed at source auto- or cross-term locations on the TF plane, respectively. A key issue in applying these methods is 'point selection', i.e. blindly choosing appropriate auto- and cross- source locations on the time-frequency plane which are suitable for applying the JD/OD criterion.

A number of approaches to TF point selection have been proposed in the literature. In the initial work of Amin *et al* [1], a criterion was suggested for discriminating peaks of the sensor STFD as belonging to auto- or cross-source distributions. One major disadvantage of this approach, is that it requires an *ad hoc* threshold to be chosen when making a decision. The appropriate value of this threshold is heavily signal dependent, as discussed in [59], where an alternative method for choosing single auto-term (SAT) locations was proposed. SAT locations are TF points at which the energy of a single source is dominant. This approach then provides information for JD only and does not exploit the cross-source terms in the separation process.

Other work on point selection has been based on hypothesis testing. In such methods, one tries to formulate a test statistic which contains information about the auto- or cross-source terms and to set an appropriate threshold based on the probability distribution function of the statistic. Initial work was based on the criterion of Amin *et al* [1]. Because the distribution of the statistic was too complicated to obtain analytically, a resampling scheme was proposed to estimate this distribution [44]. This approach, however, implies a significant computational burden. In later work, simpler test statistics were proposed in [45, 67], where the distribution could be well approximated analytically. These studies, however, did not consider the multiplicity of the problem, i.e. that we wish to test many TF locations simultaneously in the search for the desired TF points.

In this chapter, we propose a point selection scheme based on multiple hypothesis testing. The proposed approach is used to determine both auto- and cross-terms to be used in JD and/or JOD respectively. Use of a statistical MHT approach allows one to control the probability of selecting false TF locations and provides an appropriate framework for testing multiple locations on the TF plane simultaneously. Suitable statistics are proposed for determining the auto- and cross-source locations, along with second-order analysis needed to define pivotal test statistics. The chapter is organized as follows: The proposed TF point selection scheme is described in Section 4.2. The application of the proposed point selection procedure to the BSS problem is considered in Section 4.3. Conclusions are given in Section 4.4.

4.2 Testing in the time-frequency plane

Let us assume that the STFD matrix is evaluated for a set $\mathcal{S} = \{\boldsymbol{\zeta}_i; i \in [1, p]\}$ of points on the TF plane. We define the set $\mathcal{S}_A \subseteq \mathcal{S}$ of auto-source points as

$$\mathcal{S}_A \triangleq \left\{ \boldsymbol{\zeta}_k;\ \forall k \in [1, p] \text{ where } \sum_{i=1}^{d} |D_{s_i s_i}(\boldsymbol{\zeta}_k)| > 0 \right\}$$

and the set $\mathcal{S}_C \subseteq \mathcal{S}$ of cross-source points as

$$\mathcal{S}_C \triangleq \left\{ \boldsymbol{\zeta}_k;\ \forall k \in [1, p] \text{ where } \sum_{i=1}^{d-1} \sum_{j=i+1}^{d} |D_{s_i s_j}(\boldsymbol{\zeta}_k)| > 0 \right\}.$$

In practice $\mathcal{S}_A$ and $\mathcal{S}_C$ are not necessarily mutually exclusive, as the main- and sidelobes of the auto- and cross-distributions may significantly overlap.

The problem of point selection, i.e. estimating $\mathcal{S}_A$ and $\mathcal{S}_C$, is tackled herein using a hypothesis testing framework. For a single point, $\boldsymbol{\zeta}$, on the TF plane, we may wish to test the hypothesis H against the alternative K, defined respectively by

$$\begin{aligned} \mathsf{H}: &\quad g\left(\boldsymbol{D}_{\boldsymbol{ss}}(\boldsymbol{\zeta})\right) = 0, \\ \mathsf{K}: &\quad g\left(\boldsymbol{D}_{\boldsymbol{ss}}(\boldsymbol{\zeta})\right) \neq 0, \end{aligned} \tag{4.1}$$

where $g(\cdot)$ yields some combination of the source signal auto- and/or cross-TFDs at the TF point $\boldsymbol{\zeta}$. The hypothesis test is evaluated such that the probability of rejecting H when H is true is kept below a certain nominal value, termed the level of significance (LOS), while the power of the test to reject H when K is true, is maximized [82].

In the more general case, we wish to test a set, $\mathcal{S} = \{\boldsymbol{\zeta}_i; i \in [1, p]\}$ of TF locations, with the ith null and alternative hypothesis for each $\boldsymbol{\zeta}_i \in \mathcal{S}$ given respectively by

$$\begin{aligned} \mathsf{H}_i: &\quad g\left(\boldsymbol{D}_{\boldsymbol{ss}}(\boldsymbol{\zeta}_i)\right) = 0, \\ \mathsf{K}_i: &\quad g\left(\boldsymbol{D}_{\boldsymbol{ss}}(\boldsymbol{\zeta}_i)\right) \neq 0, \end{aligned} \tag{4.2}$$

for $i = 1, \ldots, p$. In this case, evaluation of the multiple hypothesis test (MHT) will result in r rejected hypotheses, of which r_{H} were true (erroneously rejected) and r_{K} were false (correctly rejected). Various approaches exist for evaluating an MHT which aim to control a global LOS, while achieving a high power [71].

Often in multiple hypothesis tests, the global level of significance, denoted α, is defined as the family wise error rate (FWE) which is the probability of rejecting at least one hypothesis when all are true:

$$\alpha = \mathsf{P}\left[\text{reject at least one } H_i | \text{all } H_i \text{ are true}\right]. \tag{4.3}$$

Strong control of the FWE implies that Equation (4.3) is satisfied for all subsets of hypotheses including the global null and therefore $\alpha \equiv \mathsf{P}\left[r_{\mathsf{H}} \geq 1\right]$. Testing procedures which strongly control the FWE include the sequentially rejective Bonferroni (SRB) tests of Holm [72] and Hochberg [70], and the closed test procedure of Hommel [73]. Although it can be shown that Hommel's test is the most powerful among these methods [74], all tests tend to give similar performance as the number of hypotheses increases. However, the complexity of Hommel's test is also the greatest. Among the relatively simple SRB procedures, Hochberg's approach is most powerful and will be used herein.

Strong control of the FWE tends to be conservative as the number of hypotheses increases. More recently, Benjamini and Hochberg have proposed controlling the expected proportion of falsely rejected null hypotheses [23]. In this case, the global level of significance, α, is termed the false discovery rate (FDR):

$$\alpha = \mathsf{E}\left[r_{\mathsf{H}}/r\right] \text{ for } r \neq 0 \text{ and } \alpha = 0 \text{ for } r = 0. \tag{4.4}$$

The FDR is equivalent to the FWE when all null hypotheses are true, otherwise it is smaller. Therefore when control of the FDR is appropriate, there is a potential for a gain in power over methods strongly controlling the FWE.

In the following we describe suitable test statistics for detecting auto- and cross-source locations and discuss how the FWE and FDR controlling procedures apply to the problem at hand.

4.2.1 Auto-Source Test Statistic

In constructing a test for the auto-source points, the function g of Equation (4.1) is chosen to be the matrix trace operation, denoted $\mathsf{Tr}\,[\cdot]$,

$$g(\boldsymbol{D}_{\boldsymbol{ss}}(\boldsymbol{\zeta})) = \mathsf{Tr}\,[\boldsymbol{D}_{\boldsymbol{ss}}(\boldsymbol{\zeta})] = \sum_i D_{s_i s_i}(\boldsymbol{\zeta}).$$

The test is therefore to determine if the sum of auto-source terms is nonzero. If we choose a TFD kernel which always yields positive peaks at the signal auto-terms, then we have a one-sided test with the ith alternate hypothesis stated as $\mathsf{K}_i :\ g\left(\boldsymbol{D}_{\boldsymbol{ss}}(\boldsymbol{\zeta}_i)\right) > 0$; $i = 1, \ldots, p$, in Equation (4.2).

A whitened and noise compensated STFD matrix is defined as

$$\begin{aligned} \underline{\boldsymbol{D}}_{\boldsymbol{XX}}(\boldsymbol{\zeta}) &= \boldsymbol{W}\big(\boldsymbol{D}_{\boldsymbol{XX}}(\boldsymbol{\zeta}) - \sigma^2\boldsymbol{I}\big)\boldsymbol{W}^H \\ &= \boldsymbol{U}\boldsymbol{D}_{\boldsymbol{ss}}(\boldsymbol{\zeta})\boldsymbol{U}^H, \end{aligned} \tag{4.5}$$

where $\boldsymbol{W}$ is the whitening matrix discussed in Section 2.1.3 and $\boldsymbol{U} = \boldsymbol{W}\boldsymbol{A}$ is a unitary matrix. Since the matrix trace operation is invariant under a unitary transformation, then $\mathsf{Tr}\,[\underline{\boldsymbol{D}}_{XX}(\boldsymbol{\zeta})] = \mathsf{Tr}\,[\boldsymbol{D}_{\boldsymbol{ss}}(\boldsymbol{\zeta})]$. We therefore propose a test statistic as follows:

$$T_A(\boldsymbol{\zeta}) = \frac{\mathsf{Tr}\left[\boldsymbol{W}\big(\hat{\boldsymbol{D}}_{\boldsymbol{XX}}(\boldsymbol{\zeta}) - \sigma^2\boldsymbol{I}\big)\boldsymbol{W}^H\right]}{\sigma_A(\boldsymbol{\zeta})}, \tag{4.6}$$

where $\sigma_A(\boldsymbol{\zeta})$ is chosen such that $T_A(\boldsymbol{\zeta})$ has unit variance under the null hypothesis.

A derivation of the variance, $\sigma_A^2(\boldsymbol{\zeta})$, for a general class of discrete-time TFDs can be found in the appendix. In the case of the PWVD, we show that

$$\sigma_A^2(\boldsymbol{\zeta}) \approx E_w \sigma_v^2 \big(2\left\|\boldsymbol{W}\right\|^2 + \left\|\boldsymbol{W}\boldsymbol{W}^{\mathrm{H}}\right\|^2 \sigma_V^2\big), \tag{4.7}$$

where E_w denotes the energy of the PWVD window function. We note that the variance is uniform across the TF plane for the PWVD, which has also been shown for the single sensor case by Stankovic [103].

4.2.2 Cross-Source Test Statistic

In the case of cross-source points, it is intuitive to define the function g of Equation (4.2), as the sum of of the magnitude of off-diagonal elements, denoted $\mathsf{Off}\,[\cdot]$;

$$g(\boldsymbol{D_{ss}}(\boldsymbol{\zeta})) = \mathsf{Off}\,[\boldsymbol{D_{ss}}(\boldsymbol{\zeta})] = \sum_{i \neq j} |D_{s_i s_j}(\boldsymbol{\zeta})|. \tag{4.8}$$

Typically the cross-terms peaks are highly oscillatory in nature and take on both positive and negative values. By considering the magnitude of components in $\boldsymbol{D_{ss}}(\boldsymbol{\zeta})$, we ensure that terms in the summation of Equation (4.8) do not add destructively.

Unfortunately, the quantity $\mathsf{Off}\,[\boldsymbol{D_{ss}}(\boldsymbol{\zeta})]$ is not easily estimated from the observed data. It is possible, however, to determine a statistic which is zero when Equation (4.8) is zero, and greater than or equal to zero when Equation (4.8) is greater than zero. Consider in the noise-free case, the quantity

$$\begin{aligned} B_k(\boldsymbol{\zeta}) &= D_{X_k X_k}(\boldsymbol{\zeta}) - m^{-1} \sum_{i=1}^{m} D_{X_i X_i}(\boldsymbol{\zeta}) \\ &= \sum_{u=1}^{d} \sum_{v=1}^{d} \left(a_{ku} a_{kv}^{*} - m^{-1} \sum_{l=1}^{m} a_{lu} a_{lv}^{*} \right) D_{s_u s_v}(\boldsymbol{\zeta}), \end{aligned} \tag{4.9}$$

for $k = 1, \ldots, m$. The auto-source contribution in the above expression occurs when $u = v$ in the double summation, which is given by

$$\begin{aligned} B_{k|\text{auto-source}}(\boldsymbol{\zeta}) &= \sum_{u=1}^{d} \left(|a_{ku}|^2 - m^{-1} \sum_{l=1}^{m} |a_{lu}|^2 \right) D_{s_u s_u}(\boldsymbol{\zeta}) \\ &= 0 \ \text{ when } |a_{ij}|^2 = P_j \ \ \forall \ \ i \in [1.m], \end{aligned}$$

where P_j is the received power of source j at the array. The difference of a sensor TFD and the average of sensor TFDs therefore results in complete cancellation of the auto-source terms, independent of the array geometry, provided the received power for each source is the same at each sensor. This condition is met when each sensor has the same gain, and the sources are in the far-field of the array.

Based on the Equation (4.9), we propose a cross-source test statistic as follows:

$$T_C(\boldsymbol{\zeta}) = \sum_{i=1}^{m} \frac{1}{\sigma_{Ci}^2} \left(\hat{D}_{X_i X_i}(\boldsymbol{\zeta}) - m^{-1} \mathsf{Tr} \left[\hat{\boldsymbol{D}}_{\boldsymbol{XX}}(\boldsymbol{\zeta}) \right] \right)^2, \tag{4.10}$$

where σ_{Ci}^2 is the variance of $\hat{D}_{X_i X_i}(\boldsymbol{\zeta})$, for $i = 1, \ldots, m$. The variances $\{\sigma_{Ci}^2(\boldsymbol{\zeta})\}_{i=1}^{m}$ are derived for a general class of discrete-time TFDs in the appendix. In the case of the PWVD, it is shown that:

$$\sigma_{Ci}^2 \approx E_w \sigma_V^2 \left(2\mathsf{Tr}\,[\boldsymbol{R_{XX}}]\,/m - \sigma_V^2 \right); \quad i = 1, \ldots, m, \tag{4.11}$$

where $\boldsymbol{R}_{\boldsymbol{XX}}$ is the spatial covariance matrix and E_w denotes the energy of the PWVD window function.

From Equation (4.9) we can see that the scaling of cross-source terms depends on the array response. Let us assume that the sensor positions are given in polar co-ordinates by $\{r_k, \phi_k\}_{k=1}^{m}$, where r is the distance of the sensor from the origin measured in carrier wavelengths and ϕ is the angle of the sensor w.r.t the origin. The response of sensor k to source l is given by

$$a_{kl} = \sqrt{P_l} \exp\left\{-\mathrm{j}2\pi r_k \cos(\phi_k - \theta_l)\right\}, \tag{4.12}$$

where P_l and θ_l denote the received power and DOA of source l, respectively. Substituting Equation (4.12) into Equation (4.9) yields

$$B_k(\boldsymbol{\zeta}) = \sum_{u=1}^{d} \sum_{v=1}^{d} \sqrt{P_u}\sqrt{P_v} b'(k, \phi_{uv}, m) D_{s_u s_v}(\boldsymbol{\zeta})$$

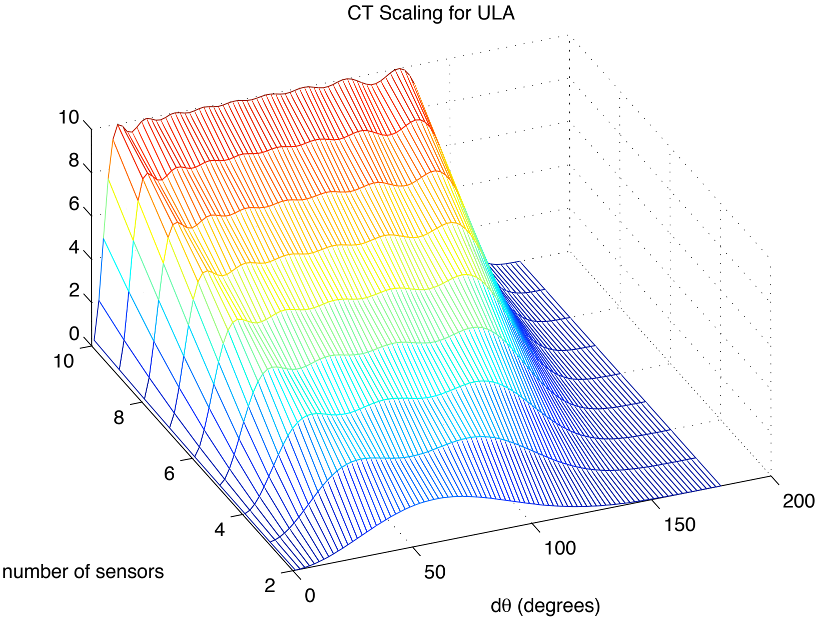
where

$$b'(k, \phi_{uv}, m) \triangleq \mathrm{e}^{\mathrm{j}4\pi r_k \sin(\phi_k)\sin(\delta\theta_{uv}/2)} - m^{-1} \sum_{l=1}^{m} \mathrm{e}^{\mathrm{j}4\pi r_l \sin(\phi_l)\sin(\delta\theta_{uv}/2)}, \tag{4.13}$$

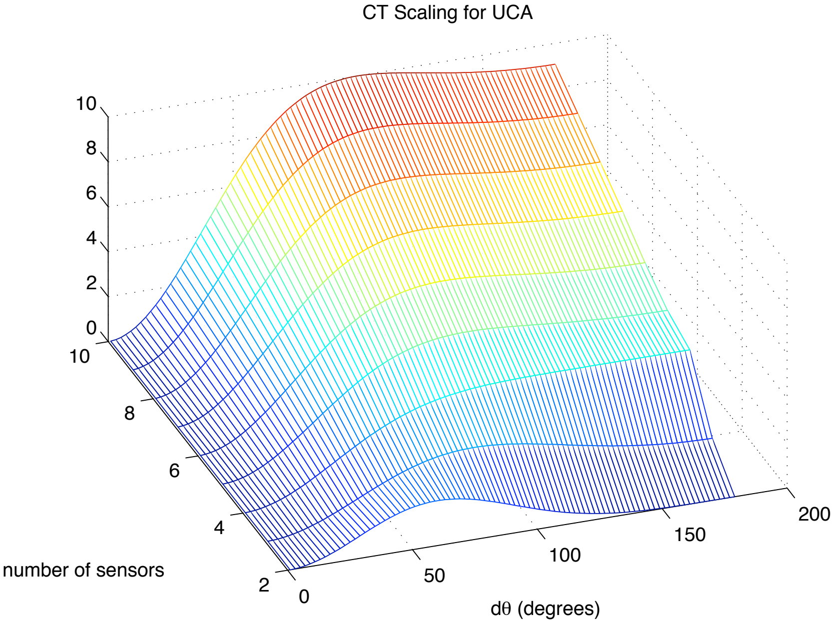
for $k = 1, \ldots, m$, where $\delta\theta_{uv} = \theta_v - \theta_u$. We note that at each sensor k, Equation (4.13) depends on the difference between the source directions, $\delta\theta_{uv}$, and the array geometry. To illustrate how cross-source terms are scaled in Equation (4.10) for two typical array configurations; a uniform linear array (ULA) and uniform circular array (UCA), we plot the function $\sum_{k=1}^{m}[b'(k, \phi_{uv}, m)]^2$ for various values of $\delta\theta_{uv}$ and m in Figure 4.1. The surface is generated, assuming the two sources have DOAs of $(-\delta\theta/2, \delta\theta/2)$ degrees broadside, by varying $\delta\theta$ from 0 to 180 degrees. We note that as $\delta\theta \to 0^\circ$ or $\delta\theta \to 180^\circ$ for the ULA and as $\delta\theta \to 0^\circ$ for the UCA, the cross-terms are scaled to zero. For other values of $\delta\theta$ the cross-terms are amplified approximately linearly w.r.t the number of sensors.

4.2.3 Evaluation of the test

To evaluate the test given in Equation (4.2), for either auto- or cross-source terms, one must know or estimate the probability distribution of the test statistic under the null hypothesis. We have found that for a PWVD window length of greater than 30 samples, the distribution of elements of $\hat{\boldsymbol{D}}_{\boldsymbol{XX}}(\boldsymbol{\zeta})$ is well approximated as multivariate Gaussian, for a range of finite variance distribution models of $\boldsymbol{V}(t)$. Under this approximation, the null distributions of $T_A(\boldsymbol{\zeta})$ and $T_C(\boldsymbol{\zeta})$ are the standard normal distribution and the Chi-Squared distribution with $m-1$



(a)



(b)

Figure 4.1: Scaling of cross-source terms for an (a) ULA and (b) UCA.

degrees of freedom (χ^2_{m-1}), respectively. Also, the use of consistent estimators of $\boldsymbol{W}$, σ_V^2 and $\boldsymbol{R_{XX}}$ instead of the true values, does not change the asymptotic distribution of the test statistics[1].

We will define the MHT procedures used in terms of p-values; $P_1, \ldots, P_p$, corresponding to hypotheses $\mathsf{H}_1, \ldots, \mathsf{H}_p$, where P_i is the lowest LOS for which H_i would be rejected based on the observed data. Let $P_{(1)} \leq P_{(2)} \leq \cdots \leq P_{(p)}$ be the ordered p-values with corresponding null hypotheses $\mathsf{H}_{(1)}, \mathsf{H}_{(2)}, \ldots, \mathsf{H}_{(p)}$. Hochberg's SRB procedure for strongly controlling the FWE is defined as [70]:

$$\begin{gathered}\text{Let } k \text{ be the largest } i \text{ for which } P_{(i)} \leq \frac{1}{p-i+1}\alpha \\ \text{then reject all } H_{(i)} \text{ for } i = 1, 2, \ldots, k.\end{gathered}$$

Benjamini and Hochberg's SRB procedure of controlling the FDR is given by [23]:

$$\begin{gathered}\text{Let } k \text{ be the largest } i \text{ for which } P_{(i)} \leq \frac{i}{p}\alpha \\ \text{then reject all } H_{(i)} \text{ for } i = 1, 2, \ldots, k.\end{gathered}$$

In the case of auto-terms, we evaluate the p-values based on Equation (4.6) and define

$$\hat{\mathcal{S}}_A = \{\boldsymbol{\zeta}_k;\ \forall k \in [1, p] \text{ where } \mathsf{H}_k \text{ is rejected}\}. \tag{4.14}$$

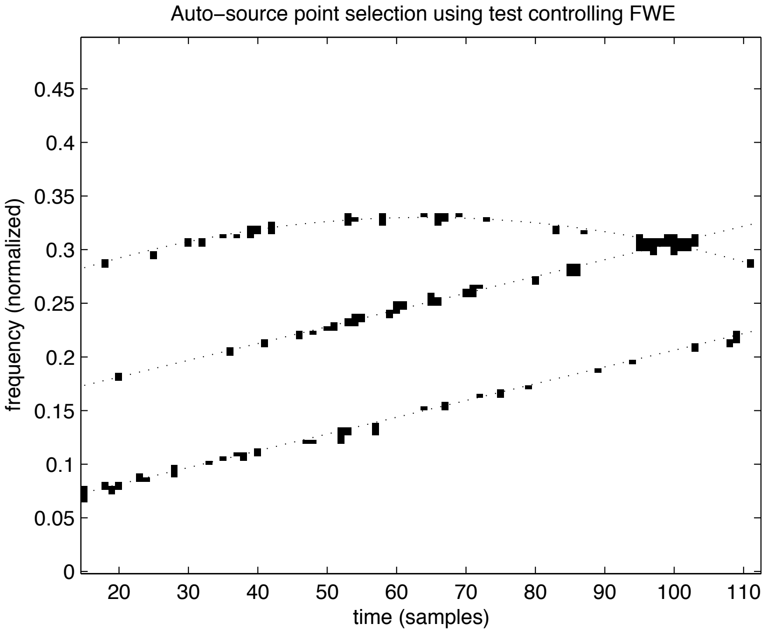
We shall denote the estimate from Equation (4.14) obtained using the FWE based procedure as $\hat{\mathcal{S}}_{A1}$ and using the FDR based procedure as $\hat{\mathcal{S}}_{A2}$. In the same way, calculating p-values based on Equation (4.10), we obtain estimates of the cross-term locations; $\hat{\mathcal{S}}_{C1}$ and $\hat{\mathcal{S}}_{C2}$. To illustrate the difference between controlling the FWE and FDR in TF point selection, we provide an example in Figure 4.2, for selection of auto-source locations. The true peaks of the auto-source distributions lie along the dotted lines. It is clear in comparing Figure 4.2 (a) and (b) that the FWE procedure has a lower number of correct detections, whereas the FDR procedure results in a larger number of false detections.

4.3 Application to blind source separation

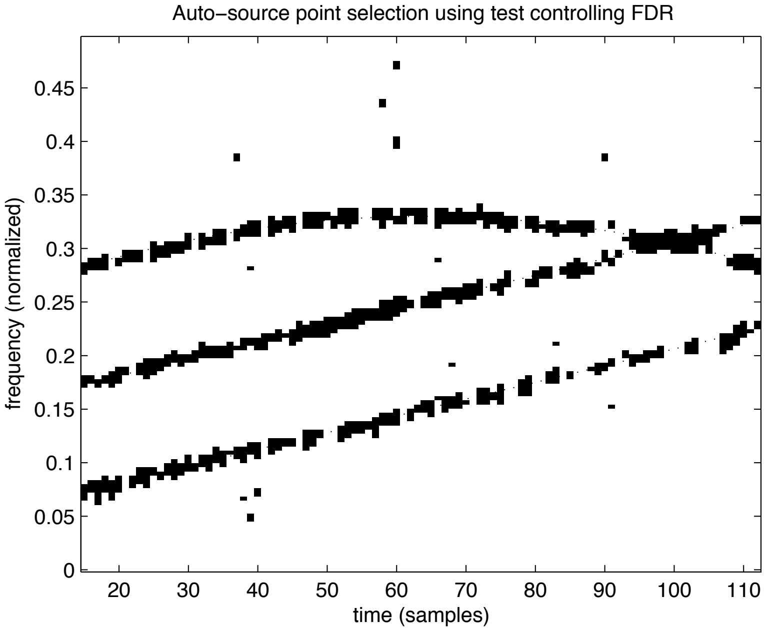
As outlined in the previous chapter, one may perform BSS through the JD/JOD of STFD matrices. The matrices to be jointly diagonalized or off-diagonalized must be constructed at TF points where $\boldsymbol{D_{ss}}(\boldsymbol{\zeta})$ has a strong diagonal or off-diagonal

---

[1] An alternative to asymptotic distributions is to use resampling techniques such as the Bootstrap [53] to estimate the null distribution. This topic is discussed further in Appendix 4.B.



(a)



(b)

Figure 4.2: Example of the MHT procedure for finding auto-term locations, using (a) the FWE and (b) the FDR controlling procedures.

structure, respectively. In order to ensure the desired structure of $\boldsymbol{D_{ss}}(\boldsymbol{\zeta})$, we consider the *dominant* auto- and cross-source points, given respectively by

$$\begin{aligned} \mathcal{S}_{\tilde{A}} &= \mathcal{S}_A \bigcap \overline{\mathcal{S}_A \bigcap \mathcal{S}_C} \qquad (4.15) \\ \text{and } \mathcal{S}_{\tilde{C}} &= \mathcal{S}_C \bigcap \overline{\mathcal{S}_A \bigcap \mathcal{S}_C}. \qquad (4.16) \end{aligned}$$

To maximize the number of points in $\mathcal{S}_{\tilde{A}}$ and $\mathcal{S}_{\tilde{C}}$, we choose a distribution characterized by φ such that the auto- and cross-term points for the signal class of interest, are both well separated and highly localized in the TF plane. An example of such a kernel is the Wigner-Ville distribution (WVD) when dealing with chirp signals. In the following, we shall make use of the PWVD, since other distributions diminish the amplitude and localization of the cross-terms which in turn degrades the performance of the JD/OD algorithm.

As discussed in the previous section, strong control of the FWE is a stricter condition than control of the FDR and therefore results in selection of only the stronger auto- or cross-source points. In the use of JD for BSS, we only need to find a few TF locations where the source STFD has strong diagonal structure (large auto-source terms). However, even the presence of weak off-diagonal terms in the source STFD can destroy the desired structure for JD. This concept holds in reverse for JOD. We therefore propose estimation of the desired TF points for BSS according to:

$$\hat{\mathcal{S}}_{\tilde{A}} = \hat{\mathcal{S}}_{A1} \cap \overline{\hat{\mathcal{S}}_{C2}} \quad \text{and} \quad \hat{\mathcal{S}}_{\tilde{C}} = \hat{\mathcal{S}}_{C1} \cap \overline{\hat{\mathcal{S}}_{A2}}, \qquad (4.17)$$

such that the strongest auto-source (resp. cross-source) terms are chosen for JD (resp. JOD) and locations with any detectable cross-source (resp. auto-source) contribution are rejected.

4.3.1 Simulation Results

To evaluate the performance of the point selection scheme, in the application to BSS, we use a variation of the performance index proposed in [17]. The power of interference source q in the pth separated source, may be expressed as $I_{pq} = \mathsf{E}|[\hat{\boldsymbol{U}}^H \hat{\boldsymbol{W}} \boldsymbol{A}]_{pq}|^2$. Overall performance is evaluated via the mean rejection level;

$$I_{perf} = (d^2 - d)^{-1} \sum_{p \neq q = 1}^{d} I_{pq}, \qquad (4.18)$$

which gives the average power of an unwanted source component in a separated source. We assess the performance in terms of the SNR and the number, K, of

TF points selected. For the proposed scheme, the LOS is set to $\alpha = 0.01$ for both the FWE and FDR controlling procedures. We then choose the $K/2$ points each from $\hat{\mathcal{S}}_{\tilde{A}}$ and $\hat{\mathcal{S}}_{\tilde{C}}$ having the largest values of T_A and T_C respectively.

We also compare the performance of two other existing approaches as follows: Belouchrani *et al* proposed classifying peaks of the sensor TFDs as auto- or cross-terms based on the criterion[2] [18]:

$$C_1(\boldsymbol{\zeta}) = \frac{\mathsf{Tr}\left[\underline{\hat{\boldsymbol{D}}}_{xx}(\boldsymbol{\zeta})\right]}{\left\|\underline{\hat{\boldsymbol{D}}}_{xx}(\boldsymbol{\zeta})\right\|}. \tag{4.19}$$

Large (resp. small) values of $C_1(\boldsymbol{\zeta})$ indicate diagonal (resp. off-diagonal) structure. However, one must set an arbitrary threshold for point selection using Equation (4.19). In this experiment, we decide a point, $\boldsymbol{\zeta}$, belongs to $\mathcal{S}_{\tilde{A}}$ or $\mathcal{S}_{\tilde{C}}$ if $C_1(\boldsymbol{\zeta}) > 0.8$ or $|C_1(\boldsymbol{\zeta})| < 0.2$ respectively[3]. We then take $K/2$ points from each set, corresponding to the largest values of $\left\|\underline{\hat{\boldsymbol{D}}}_{xx}(\boldsymbol{\zeta})\right\|$, for use in the JD/OD criterion.

A second approach, proposed by Févotte *et al* in [59], is to choose the TF locations where only a single diagonal entry of $\boldsymbol{D}_{\boldsymbol{ss}}(\boldsymbol{\zeta})$ is significant (termed SAT locations). Denoting the sample eigenvalues of $\underline{\hat{\boldsymbol{D}}}_{xx}(\boldsymbol{\zeta})$ by $\{\lambda_k^{\boldsymbol{\zeta}};\, k = 1, \ldots, d\}$, the SAT criterion is defined as

$$C_2(\boldsymbol{\zeta}) = \frac{\max_k \left|\lambda_k^{\boldsymbol{\zeta}}\right|}{\sum_{k=1}^{d} \left|\lambda_k^{\boldsymbol{\zeta}}\right|}, \tag{4.20}$$

which should be close to one at true SAT locations and smaller otherwise. We note that computation of $C_2(\boldsymbol{\zeta})$ requires an Eigen-decomposition at many points in the TF plane which implies a high computational load, whereas the p-values required for the proposed scheme can be pre-calculated and stored as a table in memory. It was suggested to choose TF points corresponding to local maxima [59], or simply the largest values [56] of the criterion in Equation (4.20). However, we obtained the best results by selecting K points from $\hat{\mathcal{S}}_{A2}$ corresponding to the largest values of Equation (4.20). This approach uses only JD to estimate the unitary transform of Equation (4.5).

The following test sources are used in the simulation experiments, consisting

---

[2] Herein $\|\cdot\|$ denotes the Frobenius norm of a matrix.

[3] These thresholds were found to give the best results in the examples considered.

of one linear FM and two quadratic FM signals:

$$\begin{aligned}
s_1(t) &= e^{j2\pi(0.05t+7.8125\times10^{-4}t^2)},\\
s_2(t) &= e^{j2\pi(0.15t+7.8125\times10^{-4}t^2)},\\
s_3(t) &= e^{j2\pi(0.25t+1.2500\times10^{-3}t^2-6.5104\times10^{-6}t^3)}.
\end{aligned}$$

The mixing system used corresponds to a ULA of $m = 8$ sensors and the sources have direction with respect to the array broadside of $(-15°, 10°, -5°)$. $N = 128$ observations are generated with unit sampling period and Gaussian noise is added. In calculation of the STFD matrices, we used the PWVD with a Hamming window of length 31 samples.

The different point selection procedures are illustrated in Figure 4.3, where the TF points selected by each scheme for $K = 40$ and an SNR of 10 dB (with respect to each source) are shown. We see that the scheme of Belouchrani *et al* chooses points for JD where both the auto- and cross-source signatures overlap. Since such points contain contributions from both auto- and cross-source TFDs, the performance of the JD/OD algorithm is significantly degraded. This algorithm tends to work better for non-overlapping TF signatures. Févotte *et al*'s method chooses only auto-source terms, though very few points corresponding to source 2 are chosen due to the overlapping cross-source signature. This can degrade the quality of the source 2 waveform estimate. The proposed scheme selects appropriate auto- and cross-source locations, though no auto-source points are chosen from source two. However, the selection of cross-source locations involving source two means that separation is still successful.

The overall performance of the point selection schemes, across a range of $K = 4$ to $K = 100$, and SNR=0 dB to SNR=20 dB, is shown Figure 4.4, where the mean rejection level has been estimated using 1000 Monte Carlo runs. Clearly the proposed method achieves the best BSS performance with respect to both SNR and number of TF points chosen. We note that the performance using $C_1(\boldsymbol{\zeta})$ is especially poor, due to the selection of points for JD at the intersection of the overlapping TF signatures. The method of Févotte *et al* is successful, though the best performance is not reached for much higher SNR and number of TF points than the proposed scheme. Though the total number of points chosen with each method was the same, the selection of some strong cross-source terms for use with JD/OD provides better performance than JD alone.

In a second example, we have sources $(s_2(t), s_3(t))$ with DOAs $(-5°, 5°)$ present. An example of each point selection procedure is shown in Figure 4.5 computed for $K = 40$ and an SNR of 10 dB. In this example we note that the method based on $C_2(\boldsymbol{\zeta})$ chooses points much farther from the TFD peaks than that of the proposed method. This is because the Eigen-value based criterion may have a high value, even at noise locations on the TF plane. The BSS performance index is evaluated across the same range of K and SNR as in the previous example, and the results are shown in Figure 4.6. We note that, while both the proposed scheme and that of Févotte *et al* are able to choose appropriate points for separation, the energy criterion fails. This is because there are no non-overlapping TF signatures, as in the previous example, from which correct points for JD would be chosen. In the results shown, the proposed method is seen to achieve good performance with fewer TF points than the method of Févotte *et al*, due both to selection of points closer to the TFD peaks and inclusion of cross-source locations as mentioned previously.

A third example uses two quadratic FM sources defined below, with DOAs $(-5^\circ, -1^\circ)$:

$$\begin{aligned} s_4(t) &= e^{\mathrm{j}2\pi(0.4t - 3.3594\times10^{-3}t^2 + 1.1393\times10^{-5}t^3)}, \\ s_5(t) &= e^{\mathrm{j}2\pi(0.3t - 2.2396\times10^{-3}t^2 + 7.5955\times10^{-6}t^3)}. \end{aligned}$$

In this case, the sources have very close TF signatures and close DOAs. An example of the point selection results from each scheme is shown in Figure 4.7. We note that here even the criterion $C_2(\boldsymbol{\zeta})$ chooses many points near the intersection of the TF signatures. This results because the DOAs are very close, so the Eigen-values of the STFD matrices are not so well separated, and the use of $C_2(\boldsymbol{\zeta})$ is therefore not as effective in discriminating SAT locations, especially in the presence of noise. The proposed method, however, is based on selecting peaks of the auto- or cross-source TFDs and therefore less effected by closely spaced sources. The BSS performance of all three point selection schemes is evaluated as previously and the results are plotted in Figure 4.8. We see that only the proposed method achieves reasonable separation of the sources. Both the energy based criterion and that of Févotte *et al* fail to determine correct TF point for separation due to the closely spaced TF and spatial signatures of the sources.

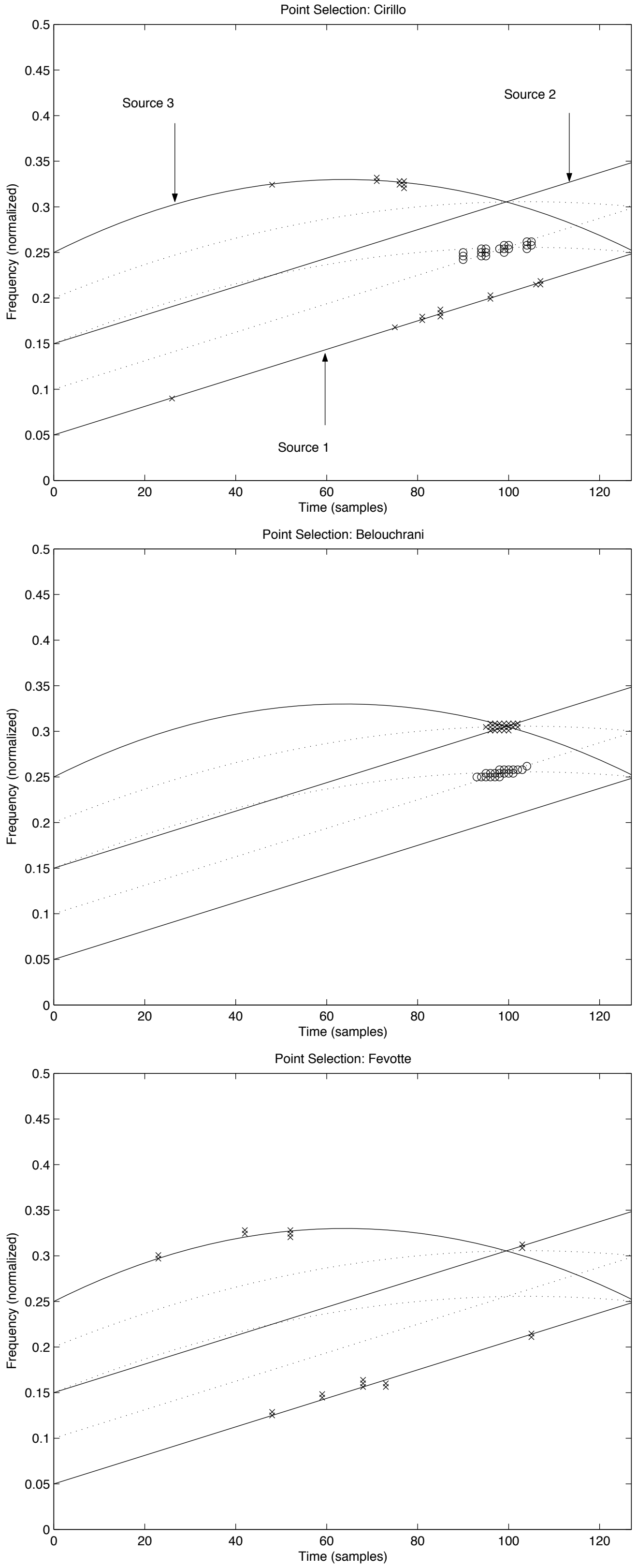
4.4 Conclusions

A ‘point selection’ scheme for selecting STFD matrices with underlying diagonal or off-diagonal structure has been proposed based on multiple hypothesis testing. The proposed method allows blind application of BSS based on JD/OD of the STFD matrices. In contrast to other proposed point selection schemes, there is no *ad hoc*, signal dependent threshold value to be chosen, rather, one decides on an acceptable probability of falsely selecting TF points. The simulation examples given here highlight a number of advantages of the proposed scheme:

- Selection of points when overlapping TF signatures are present is handled properly, even with closely spaced sources.
- By selection of both auto- and cross-source locations one can achieve better performance via JD/OD than with JD alone.
- The proposed selection scheme results in good BSS performance with fewer chosen TF points and at lower SNR than other methods.

To elaborate upon the last point: as BSS performance is achieved for a smaller number of selected points with the proposed method, fewer STFD matrices must be evaluated and the time required for optimizing the JD/OD criterion is reduced. Also, as pivotal test statistics are used, the thresholds for the MHT procedure may be pre-computed and called from a table. The computational complexity of the proposed method is therefore relatively low, compared with that of Févotte *et al*, which requires the Eigen-decomposition of many STFD matrices to compute the SAT criterion.

As a further point, one can envision the use of the proposed scheme in environments where there may be sparse sources, and given block of observations can contain no signals. In this case, the criterion based methods would still try to select points based on the largest values of the criterion, whereas the proposed method would set an appropriate threshold based on the noise power, and only (falsely) selected points according to the set LOS.


Point Selection: Cirillo

Point Selection: Belouchrani

Point Selection: Fevotte

Figure 4.3: Example of point selection based on (top) the proposed scheme (center) Belouchrani *et al* (bottom) Fèvotte *et al*. Chosen diagonal and off-diagonal terms are indicated by ‘x’ and ‘o’ respectively. The auto- and cross-source signatures are indicated by solid and dashed lines respectively.

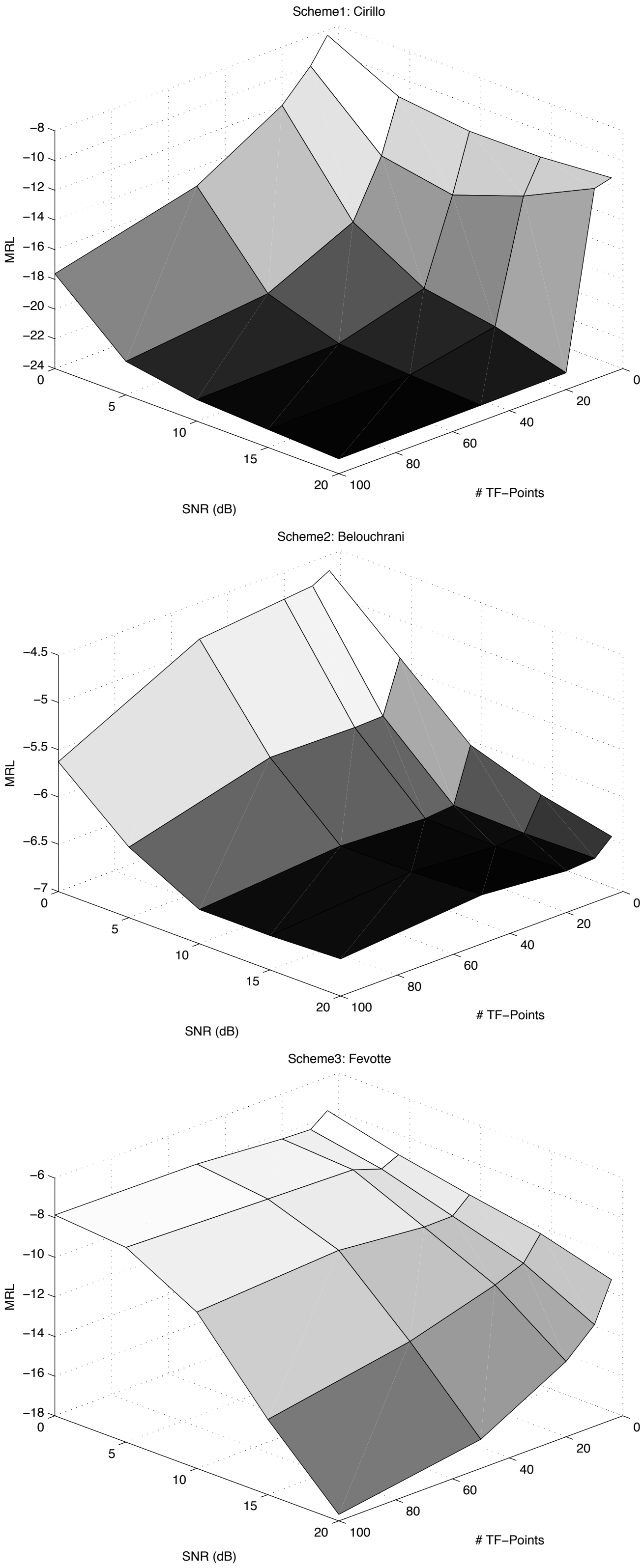


Figure 4.4: Mean rejection level versus SNR and number of TF points chosen based on (top) the proposed scheme (center) Belouchrani *et al* (bottom) Fèvotte *et al*.

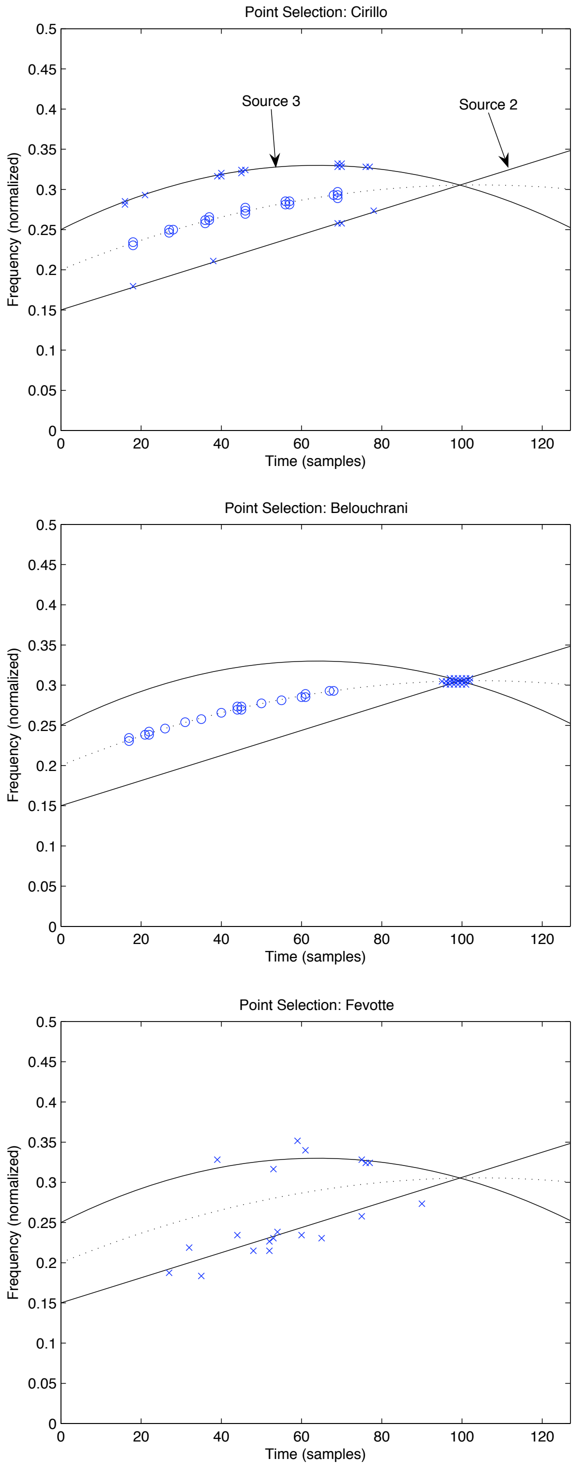


Figure 4.5: Example of point selection based on (top) the proposed scheme (center) Belouchrani *et al* (bottom) Fèvotte *et al*. Chosen diagonal and off-diagonal terms are indicated by ‘x’ and ‘o’ respectively. The auto- and cross-source signatures are indicated by solid and dashed lines respectively.

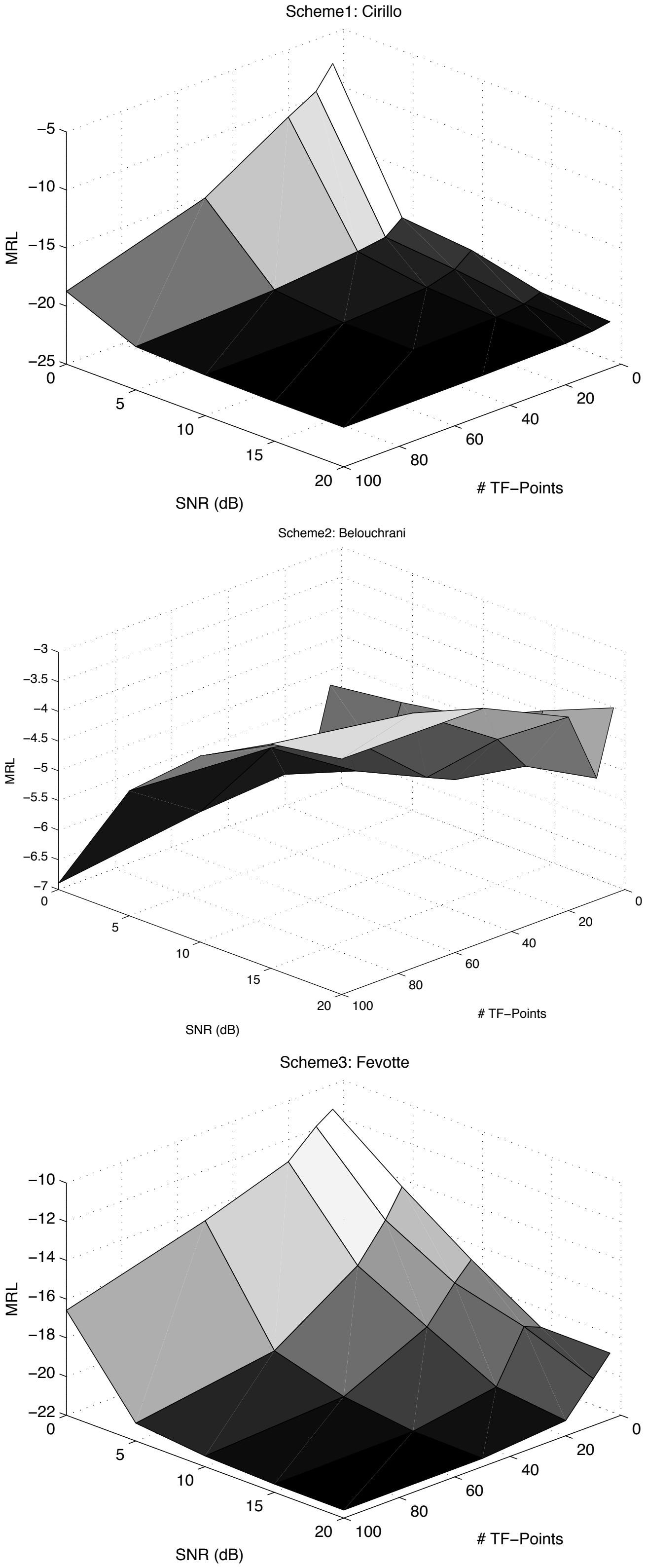


Figure 4.6: Mean rejection level versus SNR and number of TF points chosen based on (top) the proposed scheme (center) Belouchrani *et al* (bottom) Fèvotte *et al*.

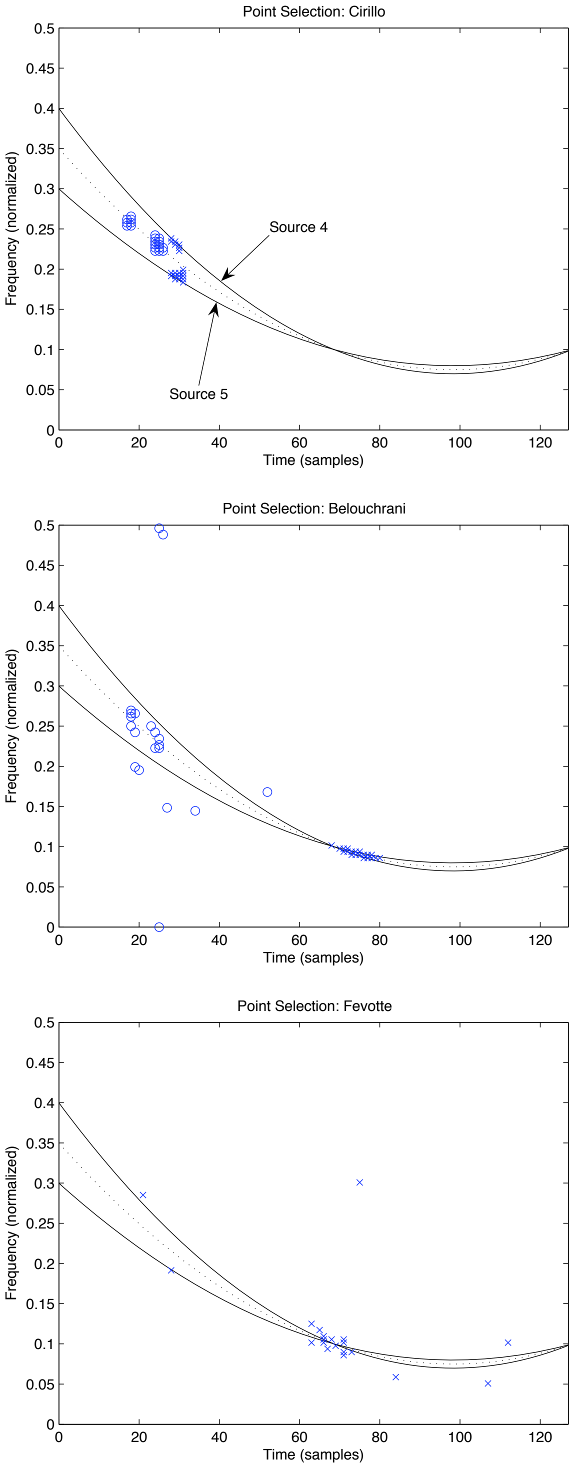


Figure 4.7: Example of point selection based on (top) the proposed scheme (center) Belouchrani *et al* (bottom) Fèvotte *et al.* Chosen diagonal and off-diagonal terms are indicated by 'x' and 'o' respectively. The auto- and cross-source signatures are indicated by solid and dashed lines respectively.

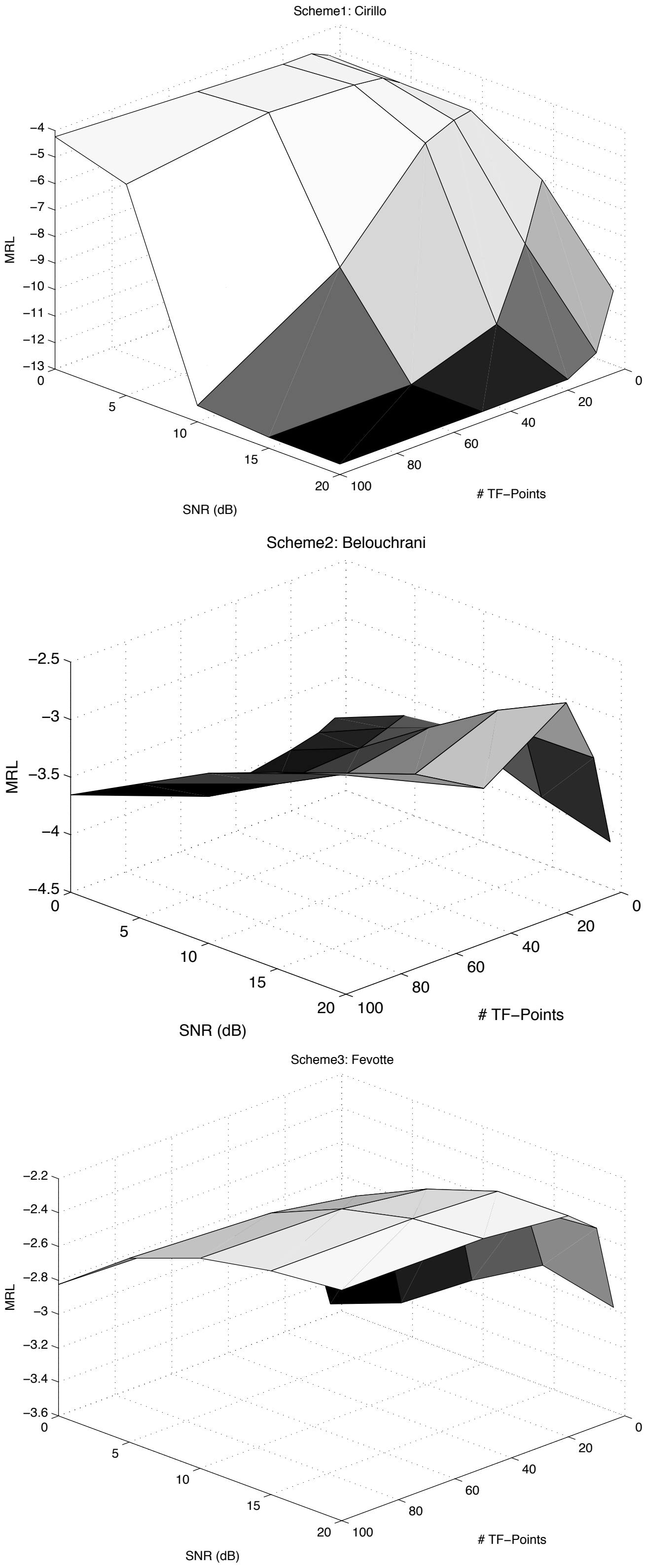


Figure 4.8: Mean rejection level versus SNR and number of TF points chosen based on (top) the proposed scheme (center) Belouchrani *et al* (bottom) Fèvotte *et al*.

4.A Derivation of the test statistic variance

In the following we derive the variances required for forming the auto- and cross-source statistics given in Equation (4.6) and (4.10) respectively. We consider the following discrete-time estimator of the STFD matrix in bilinear form:

$$\hat{\boldsymbol{D}}_{\boldsymbol{XX}}(n,\omega,\varphi) = \sum_m \sum_l \varphi(m,l)\boldsymbol{X}(n+m+l)\boldsymbol{X}^*(n+m-l)\mathrm{e}^{-\mathrm{j}2\omega l} \tag{4.21}$$

where $\varphi(m,l)$ is the kernel defining the distribution. We also require the corresponding STFD in the inner product form:

$$\hat{\boldsymbol{C}}_{\boldsymbol{XX}}(n,\omega,\psi) = \sum_m \sum_k \psi(m,k)[\boldsymbol{X}(n+m)\mathrm{e}^{-\mathrm{j}\omega m}]\ [\boldsymbol{X}(n+k)\mathrm{e}^{-\mathrm{j}\omega k}]^H \tag{4.22}$$

where $\psi(m,k) = \varphi((m+k)/2,(m-k)/2)$. We define the signal part of the observations as $\boldsymbol{y}(n) = \boldsymbol{A}\boldsymbol{s}(n)$. The STFD matrix estimator has the form

$$\hat{\boldsymbol{D}}_{\boldsymbol{XX}}(n,\omega,\varphi) = \boldsymbol{A}+\boldsymbol{B}+\boldsymbol{C}$$

where the matrices $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ are defined as

$$\begin{aligned}
\boldsymbol{A} &= \hat{\boldsymbol{D}}_{yy}(n,\omega,\varphi)\\
\boldsymbol{B} &= \hat{\boldsymbol{D}}_{yV}(n,\omega,\varphi)+\hat{\boldsymbol{D}}_{Vy}(n,\omega,\varphi)\\
\boldsymbol{C} &= \hat{\boldsymbol{D}}_{VV}(n,\omega,\varphi)
\end{aligned}$$

and the dependence on n, ω and φ is omitted from the notation for convenience. The matrices $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ shall be referred to as the *signal*, *signal-noise* and *noise* STFDs respectively. We note that only matrices $\boldsymbol{B}$ and $\boldsymbol{C}$ are comprised of random entries due to the noise.

Lemma 1 (Signal-noise STFD)

1. *The expected value is given by* $\mathsf{E}[\boldsymbol{B}] = \boldsymbol{0}$ *and the covariance of two elements of* $\boldsymbol{B}$ *is given by*

$$\mathsf{Cov}[B_{ij},B_{kl}] = \sigma_V^2\left[\hat{C}_{y_i y_k}(n,\omega,\phi)\delta_{j-l}+\hat{C}_{y_l y_j}(n,\omega,\phi)\delta_{k-i}\right] \tag{4.23}$$

where the kernel ϕ *is related to* φ *according to*

$$\phi(m_1,m_2) = \sum_k \varphi\left(\frac{m_1+k}{2},\frac{m_1-k}{2}\right)\varphi\left(\frac{m_2+k}{2},\frac{m_2-k}{2}\right). \tag{4.24}$$

This means that two elements on the same row or column of $\boldsymbol{B}$ *are correlated.*

2. *The expected value of the trace of the whitened STFD is* $\mathsf{E}\left[\mathsf{Tr}\left[\boldsymbol{W}\boldsymbol{B}\boldsymbol{W}^{\mathrm{H}}\right]\right] = 0$ *and the variance of the trace of the whitened STFD is given by*

$$\mathsf{Var}\left[\mathsf{Tr}\left[\boldsymbol{W}\boldsymbol{B}\boldsymbol{W}^{\mathrm{H}}\right]\right] = 2\sigma_v^2 \sum_{j=1}^{d}\sum_{i=1}^{d} \mathsf{Re}\left[\left[(\boldsymbol{W}\boldsymbol{W}^{\mathrm{H}}) \odot \left(\boldsymbol{W}\hat{\boldsymbol{C}}_{yy}(n,\omega,\phi)\boldsymbol{W}^{\mathrm{H}}\right)\right]_{ij}\right] \tag{4.25}$$

where $\odot$ *denotes the Hadamard or element-wise matrix product.*

Proof 1 1. *We note that* $\mathsf{E}[\boldsymbol{B}] = \boldsymbol{0}$ *since* $\mathsf{E}[\boldsymbol{V}(n)] = 0$ *and the covariance between two elements of* $\boldsymbol{B}$ *is given by*

$$\begin{aligned}\mathsf{Cov}\left[B_{ij}, B_{kl}\right] &= \mathsf{E}\left[(\hat{D}_{y_iV_j}(n,\omega,\phi) + \hat{D}^*_{y_jV_i}(n,\omega,\phi))(\hat{D}^*_{y_kV_l}(n,\omega,\phi) + \hat{D}_{y_lV_k}(n,\omega,\phi))\right] \\ &= \mathsf{E}\left[\hat{D}_{y_iV_j}(n,\omega,\phi)\hat{D}^*_{y_kV_l}(n,\omega,\phi)\right] + \mathsf{E}\left[\hat{D}_{y_lV_k}(n,\omega,\phi)\hat{D}^*_{y_jV_i}(n,\omega,\phi)\right] \\ &= \sigma_V^2(\hat{C}_{y_iy_k}(n,\omega,\phi)\delta_{j-l} + \hat{C}_{y_ly_j}(n,\omega,\phi)\delta_{k-i}),\end{aligned}$$

where the last step follows from the results of Stankovic [103] and the noise properties [6].

2. *Let* $K = \mathsf{Tr}\left[\boldsymbol{W}\boldsymbol{B}\boldsymbol{W}^{\mathrm{H}}\right]$. *Clearly* $\mathsf{E}[K] = 0$ *since* $\mathsf{E}[\boldsymbol{B}] = \boldsymbol{0}$. *The variance of* K *is therefore*

$$\begin{aligned}\mathsf{Var}[K] &= \mathsf{E}[KK^*] \\ &= \sum_{i_1=1}^{d}\sum_{i_2=1}^{d}\sum_{k_1=1}^{m}\sum_{k_2=1}^{m}\sum_{l_1=1}^{m}\sum_{l_2=1}^{m} W_{i_1k_1}W^*_{i_1l_1}W^*_{i_2k_2}W_{i_2l_2}\mathsf{Cov}\left[B_{k_1l_1}, B^*_{k_2l_2}\right] \\ &= \sigma_V^2 \sum_{i_1=1}^{d}\sum_{i_2=1}^{d}\sum_{k_1=1}^{m}\sum_{k_2=1}^{m}\sum_{l_1=1}^{m}\sum_{l_2=1}^{m} W_{i_1k_1}W^*_{i_1l_1}W^*_{i_2k_2}W_{i_2l_2} \\ &\quad \times(\hat{C}_{y_{k_1}y_{k_2}}(n,\omega,\phi)\delta_{l_1-l_2} + \hat{C}_{y_{l_2}y_{l_1}}(n,\omega,\phi)\delta_{k_1-k_2}) \\ &= \sigma_V^2 \sum_{i_1=1}^{d}\sum_{i_2=1}^{d}\Bigg\{\sum_{l=1}^{m} W_{i_2l}W^*_{i_1l}\sum_{k_1=1}^{m}\sum_{k_2=1}^{m} W_{i_1k_1}\hat{C}_{y_{k_1}y_{k_2}}(n,\omega,\phi)W^*_{i_1k_2} \\ &\quad +\sum_{k=1}^{m} W_{i_1k}W^*_{i_2k}\sum_{l_1}\sum_{l_2=1}^{m} W^*_{i_1l_1}\hat{C}_{y_{l_2}y_{l_1}}(n,\omega,\phi)W_{i_2l_2}\Bigg\} \\ &= \sigma_V^2 \sum_{i_1}^{d}\sum_{i_2=1}^{d}\Big\{[(\boldsymbol{W}\boldsymbol{W}^{\mathrm{H}}]_{i_2i_1}[\boldsymbol{W}\hat{\boldsymbol{C}}_{yy}(n,\omega,\phi)\boldsymbol{W}^{\mathrm{H}}]_{i_1i_2} \\ &\quad +[\boldsymbol{W}\boldsymbol{W}^{\mathrm{H}}]^*_{i_2i_1}[\boldsymbol{W}\hat{\boldsymbol{C}}_{yy}(n,\omega,\phi)\boldsymbol{W}^{\mathrm{H}}]^*_{i_1i_2}\Big\} \\ &= 2\sigma_v^2 \sum_{i_1}^{d}\sum_{i_2=1}^{d} \mathsf{Re}\left[\left[(\boldsymbol{W}\boldsymbol{W}^{\mathrm{H}}) \odot \left(\boldsymbol{W}\hat{\boldsymbol{C}}_{yy}(n,\omega,\phi)\boldsymbol{W}^{\mathrm{H}}\right)\right]_{i_1i_2}\right]\end{aligned}$$

Lemma 2 (Noise STFD)

1. *The expected value of the noise STFD and the covariance of two elements of $\boldsymbol{C}$ are given respectively by*

$$\begin{aligned} \mathsf{E}\left[\boldsymbol{C}\right] &= \sigma_V^2 \sum_m \varphi(m,0)\boldsymbol{I}, &(4.26)\\ \mathsf{Cov}\left[C_{ij}, C_{kl}\right] &= \sigma_V^4 \sum_m \sum_l |\varphi(m,l)|^2 \delta_{k-i}\delta_{l-j}. &(4.27) \end{aligned}$$

This results implies that any two different elements of $\boldsymbol{C}$ are uncorrelated.

2. *The expected value and the variance of the trace of the whitened noise STFD are given respectively by*

$$\begin{aligned} \mathsf{E}\left[\mathsf{Tr}\left[\boldsymbol{WCW}^{\mathrm{H}}\right]\right] &= \left\|\boldsymbol{W}\right\|^2 \sigma_V^2 \sum_m \varphi(m,0), &(4.28)\\ \mathsf{Var}\left[\mathsf{Tr}\left[\boldsymbol{WCW}^{\mathrm{H}}\right]\right] &= \left\|\boldsymbol{WW}^{\mathrm{H}}\right\|^2 \sigma_V^4 \sum_m \sum_l |\varphi(m,l)|^2. &(4.29) \end{aligned}$$

Proof 2 *1. The expected value of the noise STFD given in Equation (4.26) follows directly from the noise properties. The variance of an element $\boldsymbol{C}_{ij}$ is given by, for $i \neq j$,*

$$\mathsf{Var}\left[C_{ij}\right] = \mathsf{E}\left[C_{ij}C_{ij}^*\right] = \sum_{m_1}\sum_{l_1}\sum_{m_2}\sum_{l_2} \varphi(m_1,l_1)\varphi(m_2,l_2)^* \mathrm{e}^{-\mathrm{j}2\omega(l_1-l_2)} \cdot K$$

where

$$K = \mathsf{E}\left[V_i(n+m_1+l_1)V_i(n+m_2+l_2)\right]\mathsf{E}\left[V_j(n+m_1-l_1)^*V_j(n+m_2-l_2)\right].$$

From the noise properties, it can be seen that K is zero unless $m_1+l_1 = m_2+l_2$ and $m_1-l_1 = m_2-l_2$, or equivalently $m_1-m_2 = l_2-l_1$ and $m_1-m_2 = l_1-l_2$. This condition is only met for $m_1-m_2 = l_1-l_2 = 0$, which leads to $K = \sigma_V^4\delta_{m_1-m_2}\delta_{l_2-l_1}$ and

$$\mathsf{Var}\left[C_{ij}\right] = \sigma_V^4 \sum_m \sum_l |\varphi(m,l)|^2.$$

The variance of a diagonal element is $\mathsf{Var}\left[C_{ii}\right] = \sigma_V^4 \sum_m \sum_l |\varphi(m,l)|^2$ which follows directly from the results of Amin [6]. It is also easily verified from the noise properties that

$$\mathsf{Cov}\left[C_{ij}, C_{kl}\right] = 0 \ \forall\ (i,j) \neq (k,l).$$

2. *Let $K = \mathsf{Tr}\left[\boldsymbol{WCW}^{\mathrm{H}}\right]$, with expected value:*

$$\mathsf{E}\left[K\right] = \sigma_V^2 \sum_{m=-\infty}^{\infty} \varphi(m,0)\mathsf{Tr}\left[\boldsymbol{WW}^{\mathrm{H}}\right] = \sigma_V^2 \sum_{m=-\infty}^{\infty} \varphi(m,0)\left\|\boldsymbol{W}\right\|^2.$$

The variance of the trace of the whitened noise STFD is given by

$$\begin{aligned}
\mathsf{Var}\left[K\right] &= \sum_{i_1=1}^{d}\sum_{i_2=1}^{d}\sum_{k_1=1}^{m}\sum_{k_2=1}^{m}\sum_{l_1=1}^{m}\sum_{l_2=1}^{m} W_{i_1k_1}W^*_{i_1l_1}W^*_{i_2k_2}W_{i_2l_2}\mathsf{Cov}\left[C_{k_1l_1}, C^*_{k_2l_2}\right] \\
&= \sum_{i_1=1}^{d}\sum_{i_2=1}^{d}\sum_{k=1}^{m}\sum_{l=1}^{m} W_{i_1k}W^*_{i_1l}W^*_{i_2k}W_{i_2l}\mathsf{Var}\left[C_{kl}\right] \\
&= \sigma_V^4\sum_{m}\sum_{u}|\varphi(m,u)|^2\sum_{k=1}^{m}\sum_{l=1}^{m}\left\{[\boldsymbol{W}^{\mathrm{H}}\boldsymbol{W}]^*_{kl}\odot[\boldsymbol{W}^{\mathrm{H}}\boldsymbol{W}]^*_{kl}\right\} \\
&= \sigma_V^4\sum_{m}\sum_{u}|\varphi(m,u)|^2\left\|\boldsymbol{W}^{\mathrm{H}}\boldsymbol{W}\right\|^2.
\end{aligned}$$

We now consider the use of the PWVD which is defined by the kernel

$$\varphi(m,l) = w(l/L)\delta_m, \tag{4.30}$$

where $w(t)$ is a real positive symmetric function which is zero outside the interval $(-0.5, 0.5)$. In Equation (4.30), L is an odd integer which specifies the windowing length. Substituting Equation (4.30) into (4.24) leads to

$$\phi(m_1,m_2) = w(m_1/L)w(m_2/L)\delta_{m_1-m_2} \tag{4.31}$$

and using the above results gives

$$\hat{C}_{y_iy_j}(n,\omega,\phi) = \sum_{m=-(L-1)/2}^{(L-1)/2} w^2(m/L)y_i(n+m)y_j^*(n+m). \tag{4.32}$$

Equation (4.32) can be interpreted as a correlation between signals $y_i(n)$ and $y_j(n)$ over a window of length L centered at time n. Under the assumption of uncorrelated sources as described in Assumption 2, for a sufficient window length Equation (4.32) may be approximated as

$$\hat{C}_{y_iy_j}(n,\omega,\phi) \approx E_w\left[\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}\right]_{ij}, \tag{4.33}$$

where $E_w = \sum_n w^2(n)$ is the window function energy.

- In order for $T_A(\boldsymbol{\zeta})$ to have unit variance, we must define

  $$\sigma_A^2(\boldsymbol{\zeta}) = \mathsf{Var}\left[\mathsf{Tr}\left[\boldsymbol{W}\hat{\boldsymbol{D}}_{\boldsymbol{XX}}(\boldsymbol{\zeta})\boldsymbol{W}^H\right]\right] = \mathsf{Var}\left[\mathsf{Tr}\left[\boldsymbol{W}(\boldsymbol{B}+\boldsymbol{C})\boldsymbol{W}^H\right]\right].$$

  Therefore, by substituting Equation (4.33) into (4.25) and combining this with (4.29) we obtain the expression for $\sigma_A^2(\boldsymbol{\zeta})$ given in Equation (4.7).

- For the cross-source test statistic, $T_C(\boldsymbol{\zeta})$, we obtain (using the PWVD):

$$
\begin{aligned}
\sigma_{C_i}^2 &= \mathsf{Var}\left[\hat{D}_{X_iX_i}(\boldsymbol{\zeta})\right] = \mathsf{Var}\left[B_{ii} + C_{ii}\right] \\
&= 2\sigma_V^2 \hat{C}_{y_iy_i}(\boldsymbol{\zeta};\phi) + \sigma_V^4 \sum_m \sum_l |\varphi(m,l)|^2 \\
&= E\sigma_V^2(2[\boldsymbol{A}\boldsymbol{A}^H]_{ii} + \sigma_V^2),
\end{aligned}
$$

  for $i = 1,\ldots,m$. Assuming that each sensor has the same gain, then the diagonal elements of $\boldsymbol{A}\boldsymbol{A}^H$ are equal, while under the uncorrelated sources assumption the array covariance matrix is $\boldsymbol{R}_{\boldsymbol{X}\boldsymbol{X}} = \boldsymbol{A}\boldsymbol{A}^H + \sigma_V^2\boldsymbol{I}$, which leads to the result given in Equation (4.11).

4.B Estimation of the null distribution using Bootstrap resampling

This appendix presents an extension to the multiple hypothesis testing (MHT) framework for detecting the locations in the TF plane at which the auto-source TFDs exhibit a significant value. The MHT approach allows testing of multiple points on the TF plane simultaneously while controlling the overall probability of a false detection. *A bootstrap based scheme for estimating the null distribution of the test statistics is discussed here.* Performance of the detector using both the asymptotic and bootstrap estimates of the null distributions is presented in terms of the operating characteristics, for different SNR values.

4.B.1 Bootstrap Procedure

For each test statistic $T_i \triangleq T_A(\boldsymbol{\zeta}_i)$ defined in Equation (4.6) we require the distribution under the null hypothesis in order to evaluate significance levels for the MHT. As opposed to using an asymptotic approximation to the distribution, one may estimate the null distribution using a resampling technique known as the Bootstrap [53].

Due to the non-stationary nature of the source signals, we cannot directly resample columns from the array data matrix $\boldsymbol{X} = (\boldsymbol{x}(1),\ldots,\boldsymbol{x}(N))$, as has been done by Brcich et al for the purpose of source number detection [28]. However, the noise is assumed to be independently identically distributed across the sensors. We may therefore resample rows of $\boldsymbol{X}$ to generate a bootstrap data set $\boldsymbol{X}^*$. Recalculating the test statistic from $\boldsymbol{X}^*$ according to Equation (4.6) gives us

1. Randomly draw a set of data from a single sensor (a row of the matrix $\boldsymbol{X}$), with replacement.
2. Repeat the random selection m times to obtain a resample of the array data matrix, $\boldsymbol{X}^*$.
3. Compute the Eigen-value decomposition of the sample correlation matrix
$$\hat{\boldsymbol{R}}^* = N^{-1}\boldsymbol{X}^*(\boldsymbol{X}^*)^{\mathrm{H}}$$
and form estimates $\hat{\boldsymbol{W}}$ and $\hat{\sigma}_v^2$ of the whitening transform and noise variance respectively. [a]
4. Substitute $\hat{\boldsymbol{W}}$ and $\hat{\sigma}_V^2$ into Equation (4.7) to obtain $\hat{\sigma}_A^2$. Compute the test statistic T_i^* from Equation (4.6).
5. Repeat steps 1 to 4 B times to obtain the bootstrap test statistics $T_i^*(b)$, $b = 1, \ldots, B$.

---

[a] Estimation of $\boldsymbol{W}$ and σ_v^2 is discussed in Section 2.1.3.

Table 4.1: Bootstrap procedure for resampling non-stationary array data.

T_i^* and repeating the procedure B times then gives $T_i^*(b), b = 1, \ldots, B$ for $i = 1, \ldots, p$. The procedure is summarised in Table 4.1.

One must show caution in the procedure of Table 4.1 that the bootstrap sample correlation matrix $\hat{\boldsymbol{R}}^*$ retains at least rank $d+1$ for the estimation of the whitening transform. This problem is less likely to occur as the ratio m/d of sensors to sources becomes large. However, to ensure the rank condition is met one can add a test for the rank of $\hat{\boldsymbol{R}}^*$ to the procedure of Table 4.1, at step 3.

Since the null distribution of the test statistic is zero mean, it can be approximated using the bootstrap test statistics by forming $T_i^*(b) - B^{-1}\sum_{b=1}^{B} T_i^*(b)$, $b = 1, \ldots, B$. One can then use this distribution to determine significance levels or p-values to be used in the multiple test. In order to accurately estimate the tails of the null distribution one must have a large enough number of sensors to resample from, especially if one wishes to set a very low level of significance for the test. Experience has shown that 30 sensors are sufficient for significance levels of 0.05 and greater.

An advantage of the proposed resampling scheme is that we can estimate the null distribution in the presence of temporally correlated and non-Gaussian noise.

The only requirement is that the noise is independently distributed with respect to the sensors. Whitening methods can be considered if this scheme is to be applied in a spatially colored noise environment. An obvious disadvantage of the proposed scheme is increased computational burden since the test statistic must be computed B times instead of once.

4.B.2 Simulation examples

Under certain conditions, it can be shown that the asymptotic null distribution of the test statistic is standard Gaussian [43]. Detection based on the asymptotic Gaussian assumption has been proposed and discussed in the main body of this chapter. In the following we illustrate that in some cases the bootstrap detector can achieve better performance than detectors based on the asymptotic distribution of the test statistic.

Here we present simulation results demonstrating the performance of the MHT procedure for point selection. In the following examples, we use a uniform linear array of $m = 32$ sensors and $N = 128$ snapshots of data. There are three linear FM (chirp) source signals present, having directions of arrival (DOA) with respect to the array broadside given by $(-3^o, 0^o, 3^o)$ respectively. The noise-free sum of the chirp source signals, $\mathsf{Tr}\left[\boldsymbol{D_{ss}}(n,\omega)\right]$, is shown in Figure 4.9, where the PWVD is computed with a window length of 33. In the following we refer to the detector based on the bootstrap and the asymptotic distributions as $D1$ and $D2$ respectively.

To evaluate the performance of the point selection procedure, a set of 30 points from the TF plane is tested. We choose 15 points at which there are auto-source terms, and 15 locations which are dominated by noise. For this set of points, we apply Holm's sequentially rejective Bonferroni procedure, using both the asymptotic distribution and the bootstrap distribution of the test statistic. The number of bootstrap replications is $B = 200$. Shown in Figure 4.10 are the operating characteristic (OC) curves for SNRs of -9, -6 and -3 dB which plot the probability of detection versus the probability of false alarm.

It can be seen that both $D1$ and $D2$ have similar performance, and $D1$ slightly outperforms $D2$ for the middle SNR value. The probability of detection plotted is the overall probability of detection, i.e. the probability of rejecting *all* of the (false) H_i's. The *individual* decisions for each H_i are actually rejected at a higher rate than that shown on the OC.

Shown in Figure 4.11 is the achieved FWE rate versus the nominal level of significance of the test. It can be seen that $D2$ maintains the FWE rate at less than or equal to γ as desired. However, the error rate of $D2$ appears to be affected by the SNR value, which indicates that the asymptotic approximation of the distribution is not as accurate at lower SNR. The error rate of bootstrap approach, on the other hand, remains approximately the same regardless of the SNR value.

It can also be seen from Figure 4.11 that the bootstrap detector $D1$ deviates slightly above the set level for small values of γ. This is due to the difficulty in accurately estimating the tails of the null distribution when there is a limited number of sensors from which to resample. This problem is alleviated as the number of available sensors increases.

Both $D1$ and $D2$ exhibit a significantly lower FWE rate than the nominal level of significance for higher values of γ which indicates that the MHT procedure is conservative. However, for typically used values of γ (less than 0.1) this is not a problem. It is also possible to employ a more sophisticated MHT procedure [24], which may yield a less conservative test.

4.B.3 Summary

The idea presented in this appendix demonstrates how the application of a resampling scheme may be applied to the sensor array data, in order to estimate the null distribution of the test statistic used for point selection. This could potentially be applied to a wide range of noise models, since few assumptions on the noise distribution are made.

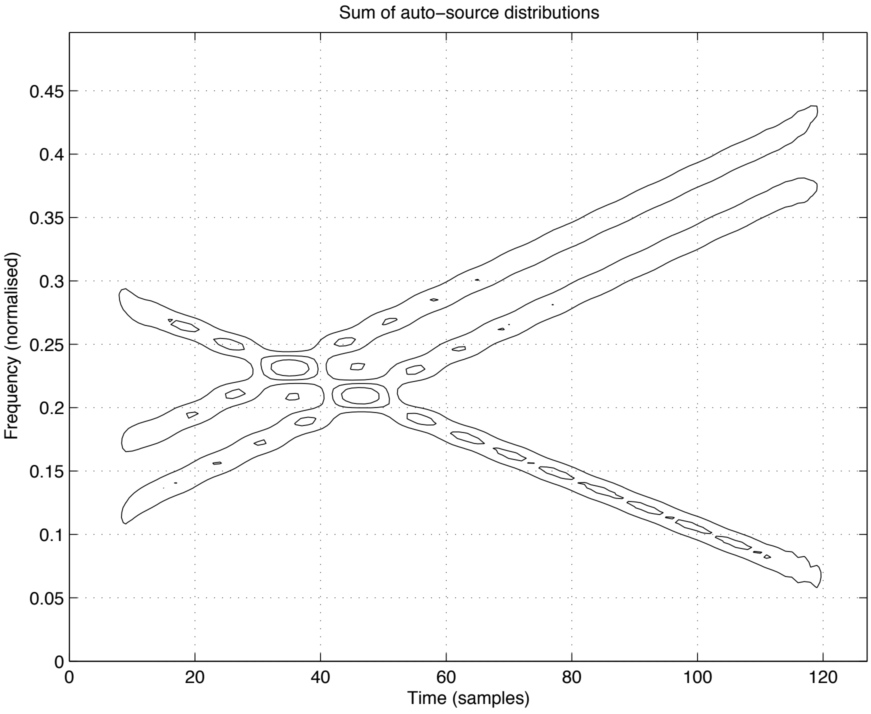


Figure 4.9: Sum of the auto-source TFDs of the three chirp source signals.

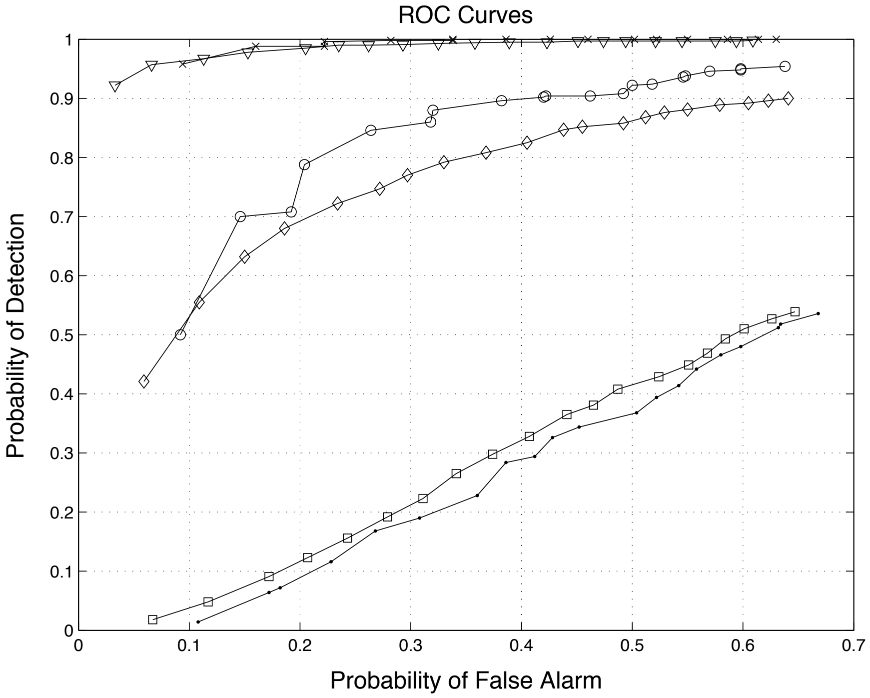
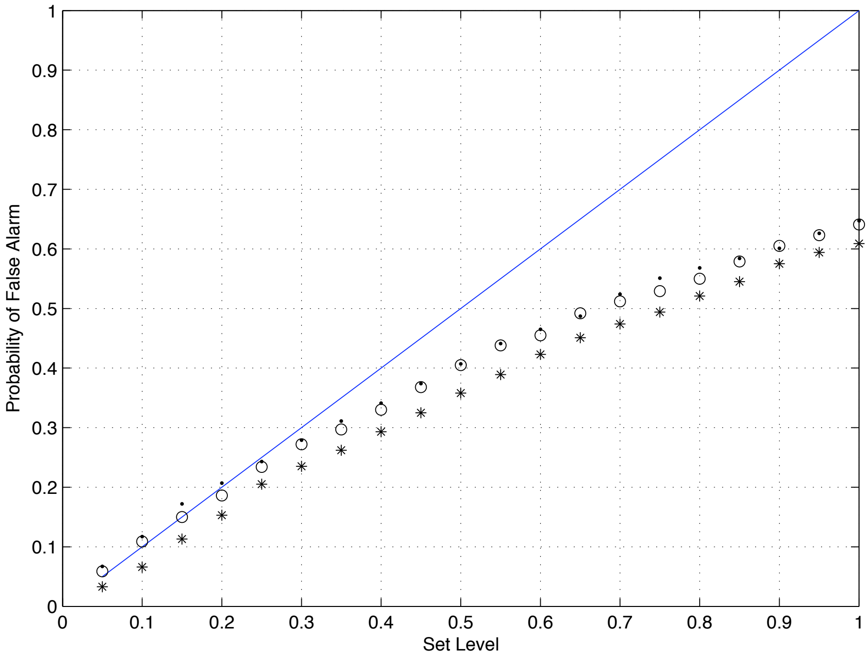


Figure 4.10: Operating Characteristic curves. SNR = -3 dB ($*$ D1, $\vee$ D2), SNR = -6 dB ($\circ$ D1, $\diamond$ D2), SNR = -9 dB ($\bullet$ D1, $\Box$ D2).



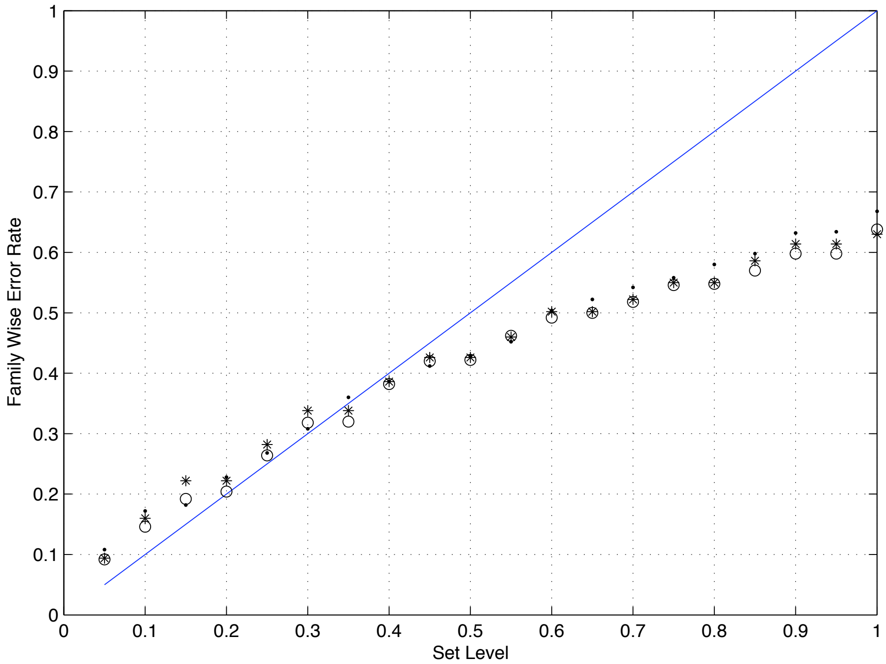


Figure 4.11: FWE for (top) normal approximation and (bottom) bootstrap based detector. SNR = -3 dB ($*$ or $\times$), SNR = -6 dB ($\circ$), SNR = -9 dB($\bullet$).

Chapter 5

Time-Frequency Signature Estimation

The problem of TF point selection was tackled in the previous chapter and demonstrated in the application of BSS. As outlined in Chapter 3, the problem of TF signature estimation is also crucial for the application of TF methods to array processing, in particular for DOA estimation. This chapter describes a parametric approach to TF signature estimation, based on the Hough transform of a TFD.

5.1 The time-frequency Hough transform

We begin with the problem of TF signature estimation based on single-sensor data. This approach is then extended to array data and applied to the DOA estimation problem in the following section.

5.1.1 Signal model

We consider the estimation of the phase parameters of mono- or multicomponent FM signals from noisy observations. The multicomponent signal model is $s(t) = \sum_{k=1}^{K} A_k e^{j\psi_k(t;\boldsymbol{\vartheta}_k)}$, where K is the number of components, $\{A_k\}$ are complex-valued amplitudes and $\{\psi(t;\boldsymbol{\vartheta}_k)\}_{k=1}^{K}$ are the phase functions parameterised by $\{\boldsymbol{\vartheta}_k\}_{k=1}^{K}$ and containing no constant term with respect to time. The IF of component k is defined according to Equation (3.9) as $\tilde{\omega}(t;\boldsymbol{\vartheta}_k) = d\psi(t;\boldsymbol{\vartheta}_k)/dt$. Given N noisy samples of $s(t)$, the problem is to estimate $\{\boldsymbol{\vartheta}_k\}$. In particular, we focus on the case of a linear FM model, which has importance in many signal processing applications.

A number of approaches exist for estimating $\{\boldsymbol{\vartheta}_k\}$ given particular phase models such as the polynomial phase transform (PPT) [94], also known as the higher-order ambiguity function (HAF), generalized product and integrated forms of the HAF [12, 13], the Wigner-Hough transform (WHT) [10, 119] and a generalized time-frequency Hough transform [11]. The WHT is of particular interest for linear FM signals, as it offers optimal detection and asymptotically efficient estimation, with an improved SNR performance threshold over other methods such as the PPT. It also provides significant suppression of cross-terms in the multicomponent case.

Statistical analysis of the WHT method was presented in [10], where the output SNR and the estimated parameter variance were derived. It was shown that the method exhibits a performance "threshold effect" in additive white Gaussian noise, such that the performance degrades rapidly for SNR below $1/N$. The method was also shown to be asymptotically efficient. However, the phase parameter estimation using the WHT requires optimization of a function with many local minima and operating in a very narrow region of attraction about the global maximum. It was suggested in [10], that this problem may be approached by first decimating and lowpass filtering the Wigner-Ville distribution (WVD) before applying the Hough transform, in order to broaden the peak centered about the true parameter values. While this approach may reduce the total number of Hough transform trajectories needed to find the global maximum, each trajectory implies a computational cost higher than that of the unfiltered WHT, which is already of order N^2.

In the following we investigate an approach based on the Hough transform of the pseudo Wigner-Ville distribution (PWVD) and demonstrate improved numerical properties with respect to the WHT based estimator. The PWVD is computed by first windowing the local auto-correlation function in the lag domain, which serves to both reduce the computational cost, and broaden the peaks of interest in the subsequent Hough transform domain, and as such, improving the numerical properties of the estimator. The price is paid in terms of a reduced SNR performance threshold and statistical efficiency with respect to the WHT. However, using the pseudo Wigner-Hough transform (PWHT) as an initialization step for the WHT provides a computationally efficient means of achieving optimal estimation, in much the same manner that nonlinear instantaneous least-squares (NLIS) improves over a direct nonlinear least-squares (NLLS) approach [8].

In the following we outline an estimation method for linear FM signals based on the PWHT. A region of attraction (RoA) for the estimator is derived which defines the required accuracy of an initial estimate or grid search, such that an efficient gradient-based optimization algorithm will converge to the true global maximum. Statistical analysis in terms of the output SNR and estimation variance is provided and compared to the corresponding results for the WHT method. In addition, we investigate the application of the PWHT to estimation of multi-component and nonlinear FM signals.

5.1.2 The Wigner-Hough Transform

The Wigner-Hough transform (WHT) of a signal $s(t)$ is defined as the line integral through the WVD of $s(t)$, along the IF model; $\tilde{\omega}(t;\boldsymbol{\vartheta})$. The WHT is therefore a mapping from the time domain to the parameter domain of $\boldsymbol{\vartheta}$. The discrete-time WHT, formed from N samples, $\{s(n)\}_{n=0}^{N}$, of a signal $s(t)$, is given by

$$W_s(\boldsymbol{\vartheta}) = \left[\sum_{n=0}^{N/2-1}\sum_{l=-n}^{n} + \sum_{n=N/2}^{N-1}\sum_{l=-(N-1-n)}^{N-1-n}\right] s(n+l)s^*(n-l)e^{-j2\tilde{\omega}(n;\boldsymbol{\vartheta})l}, \tag{5.1}$$

where N is assumed to be even. The WVD is known to be an unbiased estimator of the IF for linear FM signals [118]. However, as the IF becomes nonlinear in nature, the WHT becomes increasingly biased and in most cases, not useful. We therefore restrict our attention to a linear FM signal of the form $s(n) = Ae^{j(a_0 n + b_0/2n^2)}$, where A is a complex valued amplitude, and a_0 and b_0 correspond to the mean frequency and sweep rate respectively.

In the following, we derive expressions for the RoA of the WHT in the case of linear FM signals. Given the point (a_0, b_0) lies somewhere within the space $\Omega \subset \mathbb{R}^2$, the aim is to find a region $\Omega_c \subset \Omega$ such that a gradient-based optimization procedure initialized within Ω_c will converge to the maximum at (a_0, b_0). We derive a conservative RoA as the diamond shaped area centered at (a_0, b_0) with width and height given by the peak width in the a and b directions respectively, as illustrated in Figure 5.1. The width of the RoA in a and b will be denoted by Δ_a and Δ_b respectively.

When additive noise is present, the observations can be modelled as $X(n) = s(n) + V(n)$, $n = 0, \ldots, N-1$; where $s(n)$ is as defined previously and $V(n)$ is a complex random process. In this case, one may express the WHT of $\{X(n)\}$ as $W_X(a,b) = W_s(a,b) + \delta W_X(a,b)$. In [10], Barbarossa conducted a statistical analysis of $W_X(a,b)$ in the case of additive white Gaussian noise (AWGN) of variance σ_V^2. At the true parameter location (a_0, b_0), the variance of the perturbation term, $\delta W_X(a,b)$, was found to be

$$\mathsf{Var}\left[\delta W_X(a_0, b_0)\right] = \frac{1}{2}[N^3|A|^2\sigma_V^2 + N^2\sigma_V^4]. \tag{5.2}$$

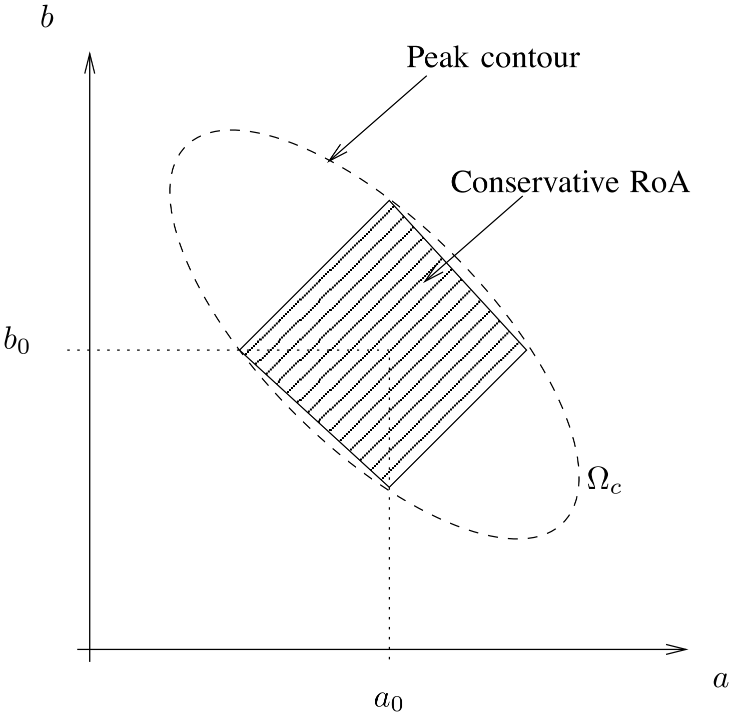


Figure 5.1: The RoA is conservatively derived as the diamond shape defined by the peak width in a and b.

In a close neighborhood of (a_0, b_0), we may therefore express the WHT as

$$W_X(a,b) = N^2[\frac{1}{N^2}W_s(a,b) + O(1/\sqrt{N})]. \tag{5.3}$$

We note that $W_s(a,b)$ takes its maximum value at (a_0, b_0) of $N^2|A|^2/2$, which is $O(N^2)$. Thus, from Equation (5.3) we observe that, to within $O(1/\sqrt{N})$, the shape of $W_x(a,b)$ is determined by the function $W_s(a,b)$, in a close neighborhood of the true parameter values. We therefore determine an approximate RoA for the WHT estimator based on the shape of $W_s(a,b)$.

It can be readily shown by direct substitution into Equation (5.1) that

$$\begin{aligned} W_s(a,b_0) &= 2\left[\frac{\sin(\bar{a}N/2)}{\sin(\bar{a})}\right]^2 && (5.4) \\ W_s(a_0,b) &= \left[\sum_{n=0}^{(N/2-1)} - \sum_{n=N/2}^{(N-1)}\right]\frac{\sin(\bar{b}n(1+2n-2N))}{\sin(\bar{b}n)} && (5.5) \end{aligned}$$

where $\bar{a} = a_0 - a$ and $\bar{b} = b_0 - b$. From Equation (5.4) it is clear that the main peak width is $\Delta_a = 2\pi/N$. However, it is not easy to analytically solve for the peak width of Equation (5.5) due to the nonlinear nature of the expression. We have numerically solved[1] for the main peak width of (5.5) across a range of values of N, as the distance between the first local minima either side of the origin ($\bar{b} = 0$). The useful approximation; $\Delta_b \approx 8\pi/[N(\frac{10}{9}N - 1)]$, was found to deviate from the numerical solution by a maximum of 1.769%, across the range of values $N = 2^3$ to $N = 2^{14}$.

Restricting attention to the case of non-aliased[2] linear FM signals, we note that $a \in [0, \pi)$ and $b \in [-a/N, (\pi - a)/N]$, i.e. we must search for the true value of a and b over intervals of π rad/s and π/N rad/s^2 respectively. From the previous discussion, we conclude that the RoA for the WHT is $O(1/N)$ in a and $O(1/N^2)$ in b, though we may say the RoA is $O(1/N)$ for both a and b with respect to the non-aliased parameter range. The problem of estimating a and b using the WHT is therefore evident; the desired peaks of the WHT become narrower as the number of observations increases, making the peak finding problem increasingly difficult.

5.1.3 The pseudo Wigner-Hough Transform

In the discrete-time case, the pseudo Wigner-Hough transform (PWHT) is calculated from N samples $\{s(n)\}_{n=0}^{N}$ of $s(t)$ as,

$$P_s(\boldsymbol{\vartheta}) = \sum_{n=M}^{(N-M-1)} \sum_{l=-M}^{M} s(n+l)s^*(n-l)e^{j2\tilde{\omega}(n;\boldsymbol{\vartheta})l} \tag{5.6}$$

where M is a parameter defining the odd PWVD window length; $L = 2M + 1$. We have defined the PWHT as the summation over the $N - L + 1$ points in the center of the PWVD, leaving out the rising and falling edges of the distribution, assuming that $L << N$. Defining $R_s(n, l) = s(n + l)s^*(n - l)$ as the local auto-correlation function, we observe that the PWHT is a weighted sum over the center rectangle of the support of $R_s(n, l)$, whereas the WHT sums over the full diamond-shaped support, as illustrated in Figure 5.2. As such the computation time of the PWHT is less than that of the WHT. The number of complex multiplications and additions required is summarised in Table 5.1 (for N even), which shows that the PWHT has reduced computational cost by a factor of approximately $2L/N$ for $L << N$.

---

[1] Optimization routines available within the software package MATLAB were successfully employed to this end.

[2] Due to the inherent frequency axis scaling of the discrete-time WVD, non-aliased signals are those whose IF lies within $[0, \pi)$ rad/s within the observation interval.

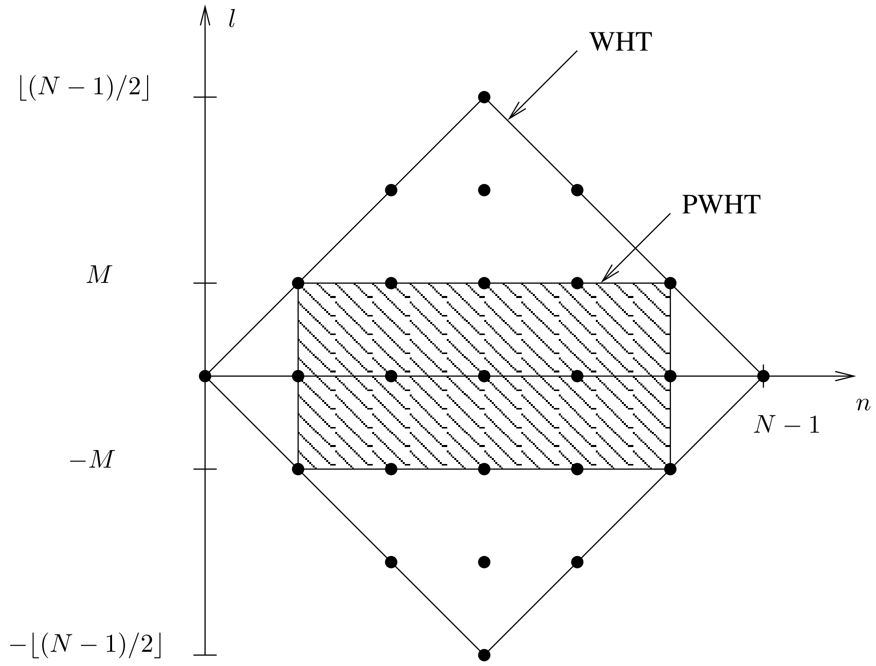
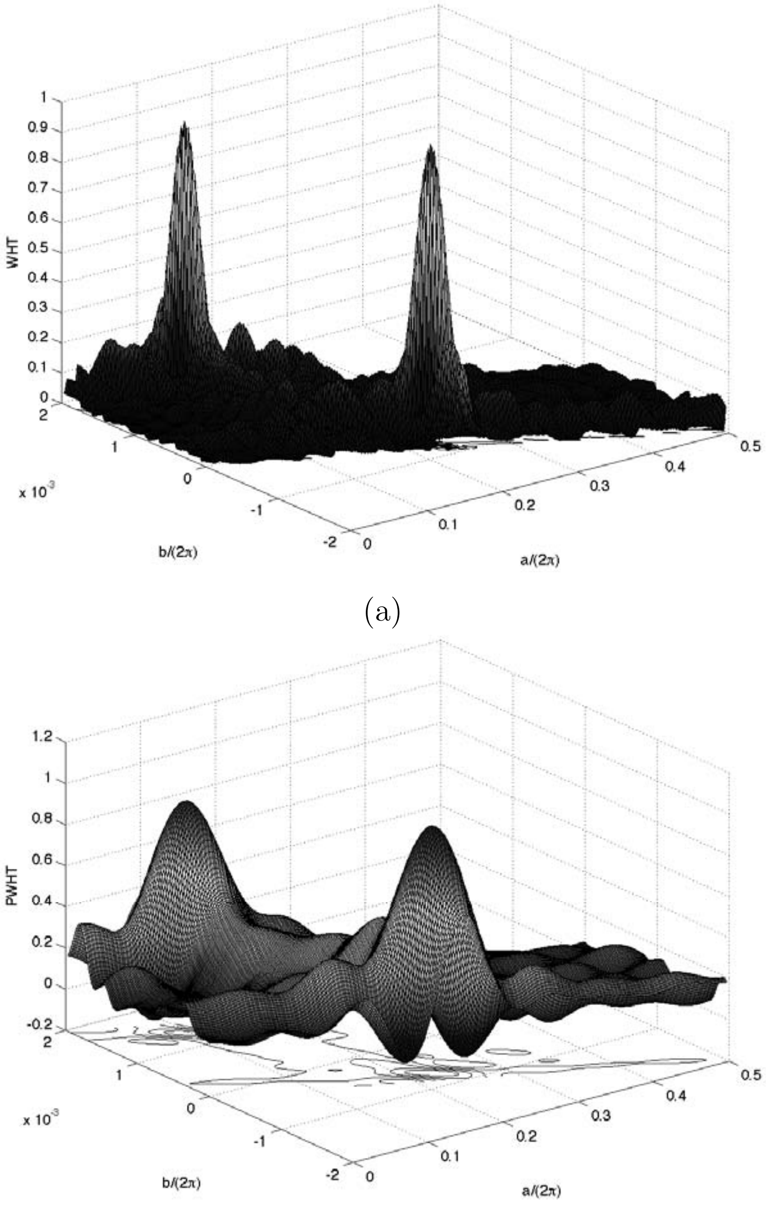


Figure 5.2: Support of $R_s(n,l)$ and the points used in computation of the PWHT and WHT.

Method	# Complex Mult's	# Complex Add's
WHT	N^2	$N^2/2-1$
PWHT	$2L(N-L+1)$	$L(N-L+1)-1$

Table 5.1: Comparison of the computational complexity of the PWHT and the WHT.

Apart from reducing the computational complexity, the PWHT has another important and fundamental advantage over the WHT. The PWVD may be seen as the WVD convolved with a sinc function in the frequency domain, which results in a widening of the main lobe of any signal component. Figure 5.3 depicts the PWHT and WHT functions for a two linear FM component signal observed in AWGN at an SNR of 0dB, with $N = 128$ and $M = 5$. Clearly the peak width in the parameter space is much larger for the PWHT than for the WHT. This implies improved numerical properties with respect to the WHT, when performing optimization, i.e. a larger RoA. Of course, there is also a disadvantage in the windowing used in the PWHT, which is also evident in Figure 5.3. Namely, the noise floor has increased due to the shorter summation interval used. This implies that the SNR performance threshold and estimator variance of the PWHT is worse than that of the WHT. In order to clarify these points, we derive in the following sections expressions for the RoA of the PWHT, the output SNR, and the statistical accuracy.



(a)

(b) (a) (b)

Figure 5.3: (a) WHT and (b) PWHT of the sum of two chirp signals.

5.1.3.1 Region of Attraction

Herein, we determine an approximate expression for the RoA of the PWHT estimator of linear FM signals. The same notation is used as previously in Section 5.1.2. In the case of noisy observations, the PWHT may be expressed as $P_x(a,b) = P_s(a,b) + \delta P_x(a,b)$. As shown in Appendix 5.A, the variance of the perturbation term $\delta P_X(a,b)$ at the true parameter location (a_0,b_0) is given by

$$\mathsf{Var}\left[\delta P_X(a_0,b_0)\right] = \frac{2}{3}\sigma_V^2|A|^2L(3NL-(L-1)(4L+1)) + \sigma_V^4(N-L+1)L, \quad (5.7)$$

which is $O(NL^2)$. In a close neighborhood of (a_0,b_0), we may therefore express the PWHT as

$$P_X(a,b) = NL[\frac{1}{NL}P_s(a,b) + O(1/\sqrt{N})]. \quad (5.8)$$

We note that $P_s(a,b)$ takes its maximum value at (a_0,b_0) of $(N-L+1)L|A|^2$, which is $O(NL)$. Accordingly, to within $O(1/\sqrt{N})$, the shape of $P_X(a,b)$ is determined by the function $P_s(a,b)$, in a close neighborhood of the true parameter values. We therefore determine an approximate RoA for the PWHT estimator based on the shape of $P_s(a,b)$.

To determine Δ_a and Δ_b, we first compute[3] the derivatives of $P_s(a,b)$ with respect to $\bar{a} = a_0 - a$ and $\bar{b} = b_0 - b$. Substituting the linear FM signal into Equation (5.6) and taking $b = b_0$ we obtain:

$$\begin{aligned} P_s(a,b_0) &= (N-2M)\frac{\sin(\bar{a}(2M+1))}{\sin(\bar{a})} \\ \frac{\partial P_s(a,b_0)}{\partial \bar{a}} &= (N-2M)\frac{[M\sin(2\bar{a}(M+1)) - (M+1)\sin(2\bar{a}M)]}{\sin^2(\bar{a})}. \end{aligned} \quad (5.9)$$

Substituting the linear FM signal into Equation (5.6) and taking $a = a_0$ we obtain:

$$\begin{aligned} P_s(a_0,b) &= \sum_{n=M}^{N-M-1}\frac{\sin(\bar{b}(2M+1)n)}{\sin(\bar{b}n)} \\ \frac{\partial P_s(a_0,b)}{\partial \bar{b}} &= \sum_{n=M}^{N-M-1}\frac{n[(M+1)\sin(2\bar{b}Mn) - M\sin(2\bar{b}(M+1)n)]}{\sin^2(\bar{b}n)}. \end{aligned} \quad (5.10)$$

To determine the peak width in a, one may solve Equation (5.9) for the location of the first zeros occurring either side of $\bar{a} = 0$. However, for Equation (5.10) this approach is not always valid. Depending on the value of M and N, the peak in $\bar{b}$ may not have local minima at the edges, as illustrated in Figure 5.4, where $P_s(a,b)$ and its derivative are plotted for $M = 10$ and $N = 128$. In this case we define the peak edges as the midpoint between the first and second points of

---

[3] We work with the window parameter $M = (L-1)/2$, as it leads to more compact expressions.

inflection. Determining Δ_a and Δ_b, therefore, requires solving for the zeros of $\partial P_s(a, b_0)/\partial \bar{a}$ and $\partial^2 P_s(a_0, b)/(\partial \bar{b})^2$ respectively, where

$$\begin{aligned}\frac{\partial^2 P_s(a_0,b)}{(\partial \bar{b})^2} = \sum_{n=M}^{N-M-1} & 2n^2 \csc^2(2\bar{b}n)\big[M(M+1)\cos(2\bar{b}Mn) \\ & -M(M+1)\cos(2\bar{b}(M+1)n) \\ & -2\cot(2bn)[(M+1)\sin(2\bar{b}Mn) \\ & -M\sin(2\bar{b}(M+1)n)]\big]. \qquad (5.11)\end{aligned}$$

For a particular value of M and N, one may numerically solve for the zeros of the nonlinear equations (5.9) and (5.11) to determine Δ_a and Δ_b. We have performed the numerical analysis across a wide range of values of M and N and found the following useful empirical approximations:

$$\Delta_a \approx \frac{2\pi}{1.4(2M+1)} = \frac{2\pi}{1.4L}, \qquad (5.12)$$

$$\begin{aligned}\Delta_b &\approx \frac{2\pi}{(-1.19M^2 + 1.2MN - 3M + 0.4N + 17.5)} \\ &= \frac{2\pi}{18.7025 - 0.2975L^2 + L(-0.905 + 0.6N) - 0.2N}. \qquad (5.13)\end{aligned}$$

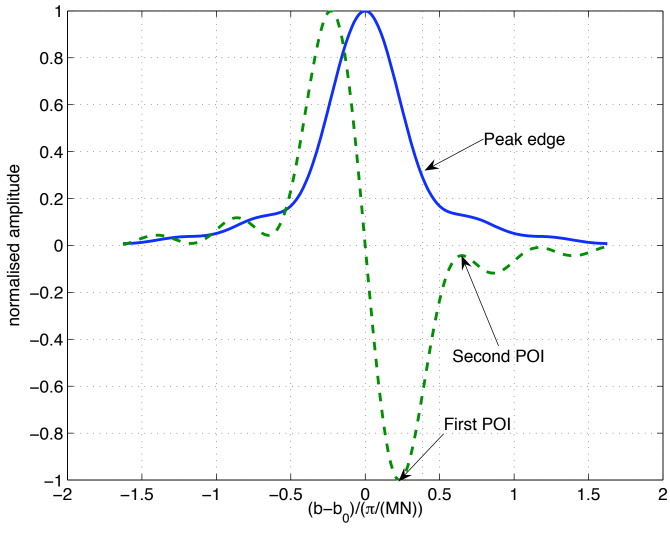


Figure 5.4: Normalized plot of $H_s(a_0, \bar{b})$ versus $\bar{b}$ (solid) and its derivative with respect to $\bar{b}$ (dashed).

The expressions (5.12) and (5.13) demonstrate that the RoA in both a and b increases with reduced window length, in the computation of the PWHT. In comparison to the ROA of the WHT, the ROA of the PWHT increases in a from $O(1/N)$ to $O(1/L)$ and in b from $O(1/N^2)$ to $O(1/(LN))$. Again, considering only the non-aliased parameter ranges, the RoA for both a and b is $O(1/L)$. This results is particularly interesting, because the RoA is now *independent of the number of samples*, and as such we may *control the RoA using the window parameter*. Further, the Equations (5.12) and (5.13) can be used to determine the appropriate grid sampling density for performing a grid search of the parameter space using the PWHT. By sampling along the a and b parameter directions with spacing $\Delta_a/2$ and $\Delta_b/2$ respectively, we can ensure that one of the trajectories evaluated will lie within the RoA of the PWHT, yielding a suitable initialization for optimization routines. The implementation of the PWHT is further elaborated in the final part of this section.

To illustrate the results derived above, the PWHT and the approximate RoA specified by Equations (5.12) and (5.13) are plotted in Figure 5.5, over the range of non-aliased parameter values. In this example, we show the PWHT function simulated at an SNR of 0 dB, for various values of N and M. The black diamond and cross in each sub-figure show the approximate RoA and true parameter locations respectively. In all sub-figures, the approximate RoA regions lie over the main peak contour as expected. Comparing Figure 5.5 (a) and (c) we observe that although the sample lengths are different ($N = 64$ and $N = 128$ respectively), using the same PWVD window length ($L = 13$) yields the same RoA. This illustrates how the RoA is independent of the sample length, with respect to the range of non-aliased parameter values. Comparing Figure 5.5 (b)-(d) we observe how increasing the PWVD window length decreases the RoA, for $L = 7$, 13 and 27 respectively.

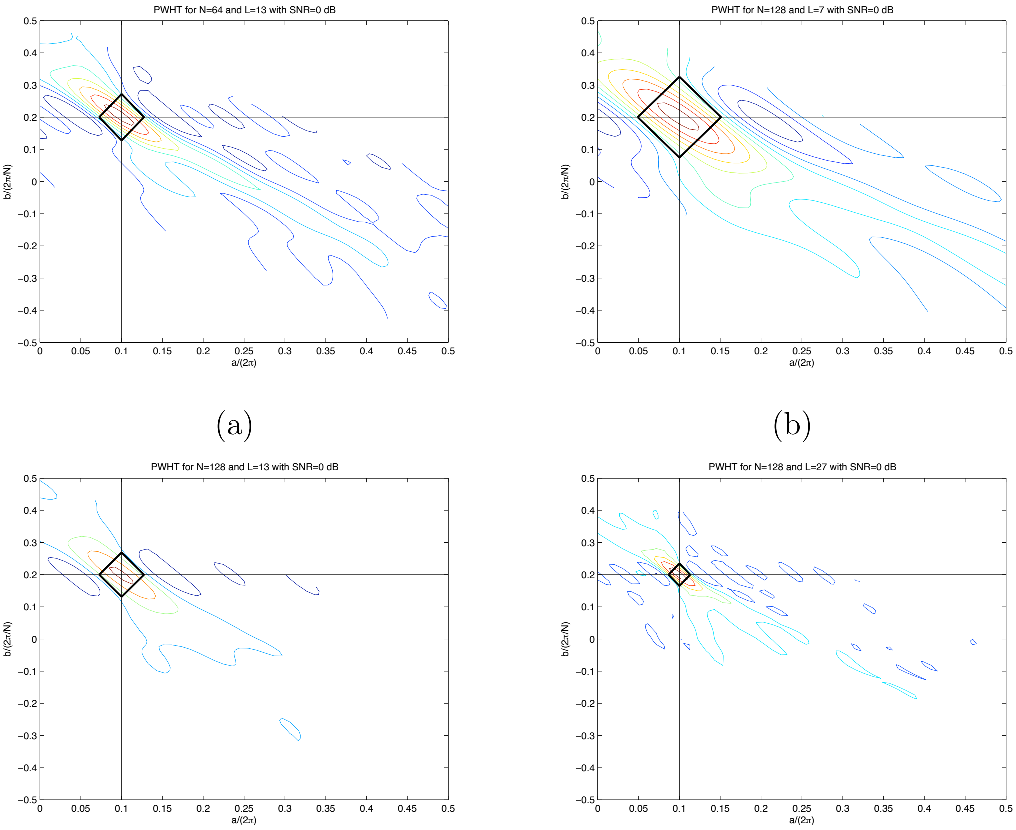
5.1.3.2 Output SNR

In the following, we perform a statistical analysis of the PWHT based estimate, in terms of the output SNR. In the case of nonlinear FM signals, the analysis holds based on the assumption:

Assumption 5 (Locally linear IF)
We assume that $\exists M \in \mathbb{Z}^+$ *and an interval* $\mathcal{I}_n = \{n - M, \ldots, n + M\}\}$ *such that:*

$$\tilde{\omega}(k; \boldsymbol{\vartheta}) \approx a_n + b_n k \quad \forall k \in \mathcal{I}_n; \text{ and } \forall n \in \{M, \ldots, N - M - 1\},$$

i.e. the signal IF is approximately linear within all time intervals of length L, over the entire observation period.



(a) (b)

(c) (d)

Figure 5.5: An example of the PWHT at SNR=0 dB, for (a) $N = 64$ and $L = 13$, (b) $N = 128$ and $L = 7$, (c) $N = 128$ and $L = 13$ and (d) $N = 128$ and $L = 27$. The approximate RoA given by Equations (5.12) and (5.13) is shown as a black diamond. The true parameter location at the intersection of the vertical and horizontal solid lines.

We also assume that the noise $V(n)$ is a complex white Gaussian process of variance σ_V^2.

Given a model for the observations $\{X(n) = s(n) + V(n)\}_{n=0}^{N-1}$ where $s(n) = Ae^{j\psi(n;\boldsymbol{\vartheta}_0)}$, the output SNR is defined as

$$SNR_{out} = \frac{|P_s(\boldsymbol{\vartheta}_0)|^2}{\mathsf{Var}\left[P_X(\boldsymbol{\vartheta}_0)\right]}. \tag{5.14}$$

Assuming the value of M is chosen such that Assumption 5 is satisfied for the given signal and using Equation (5.6), we obtain the following expression for the output SNR in terms of the input SNR; $SNR_{in} = |A|^2/\sigma_V^2$:

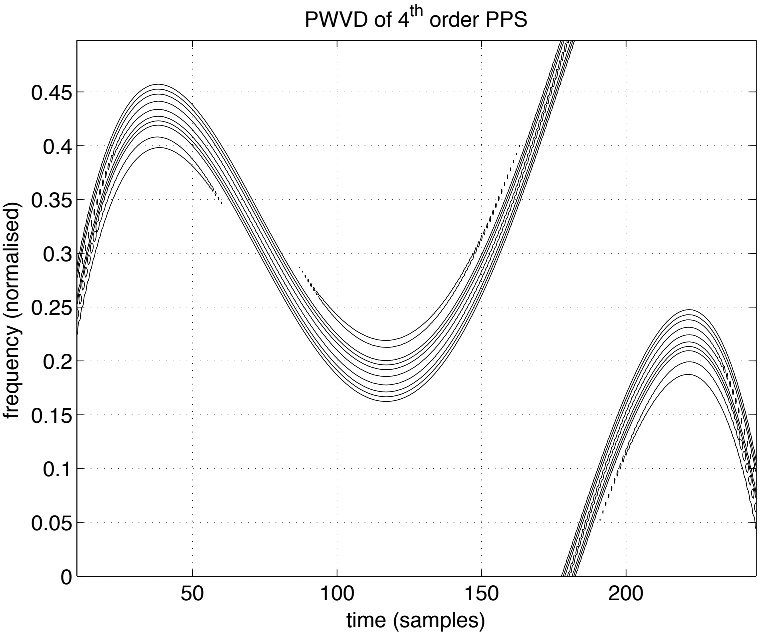
$$SNR_{out} \approx \frac{(N-L+1)L \cdot SNR_{in}^2}{\left(\frac{2}{3}\frac{3NL-(L-1)(4L+1)}{N-L+1}\right) SNR_{in} + 1}, \tag{5.15}$$

where the above expression becomes exact in the case of linear FM signals. The derivation of Equation (5.15) is included as Appendix 5.A. Assuming $L << N$, then the denominator of Equation (5.15) is approximately equal to $(2L)SNR_{in}+1$. We therefore see a thresholding effect versus the input SNR; for $SNR_{in} >> 1/(2L)$, the output SNR is approximated as $SNR_{out} \approx N(SNR_{in}/2)$. For $SNR_{in} << 1/(2L)$ the output SNR degrades rapidly according to $SNR_{out} \approx (NL)SNR_{in}^2$. The SNR performance threshold of the method is therefore said to occur at the interception point of these two limiting behaviors, i.e. at $SNR_{in} = 1/(2L)$. In comparison with the the WHT (suitable only for linear FM signals) which has an SNR performance threshold at $1/N$, the threshold of the PWHT is higher (worse), however the output SNR performance above the input SNR threshold is the same.

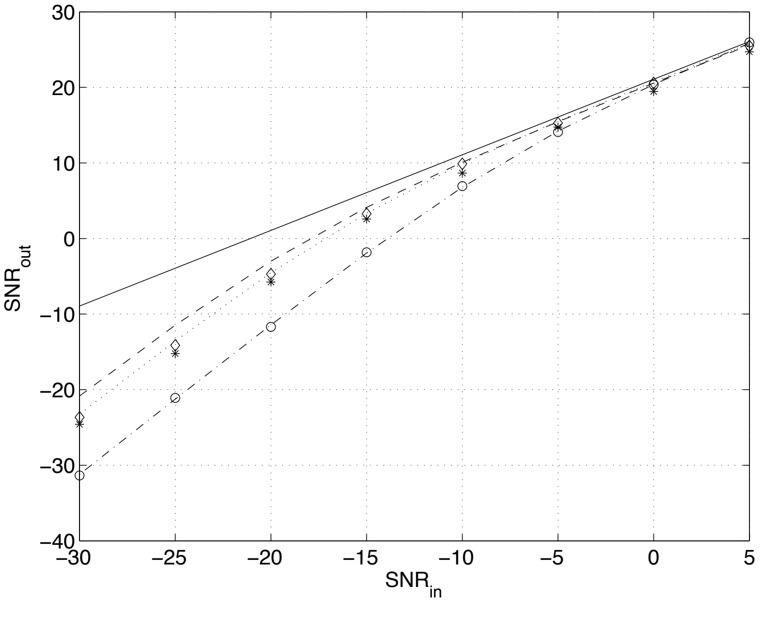
In Figure 5.6 we plot theoretical and simulated values of SNR_{out} versus SNR_{in} for various window lengths. In this example we use $N = 256$ samples of a fourth order polynomial phase signal, of which the PWVD is shown in Figure 5.6 (a) using $L = 21$. We note that for $L = 3$ and $L = 21$ the simulation shows close agreement with theory. However for $L = 41$ the IF is no longer close to linear within all summation intervals and the simulation shows an output SNR worse than that for $L = 21$. Clearly, having a larger window length is advantageous in lowering the SNR threshold. However, as L increases the estimation of nonlinear IF laws becomes increasingly biased and the approximation given in (5.15) is no longer valid, since the peak location in the noise free case moves away from $\boldsymbol{\vartheta}_0$.

5.1.3.3 Estimation accuracy

The mean and variance of the PWHT estimator are derived below via a generalization of the perturbation approach taken in [10], for analysis of the WHT. Assuming the same signal model as in the previous section, we may express the PWHT of the observations as the sum of $P_s(\boldsymbol{\vartheta})$ being the PWHT of the noise-free signal and a perturbation $\delta P(\boldsymbol{\vartheta})$ composed of cross signal-noise and noise only terms. Under Assumption 5, the maximum of $P_s(\boldsymbol{\vartheta})$ occurs at $\boldsymbol{\vartheta}_0$, but in the presence of noise, the maximum of the PWHT shifts to a location $\boldsymbol{\vartheta}_0 + \delta\boldsymbol{\vartheta}$ due to the influence of $\delta P(\boldsymbol{\vartheta})$. The bias and variance of the estimator are therefore determined by the mean and covariance matrix of $\delta\boldsymbol{\vartheta}$. As a first order approximation, the bias $\mathsf{E}\left[\delta\boldsymbol{\vartheta}\right] = 0$ and the covariance matrix $\boldsymbol{\Gamma}_{\delta\boldsymbol{\vartheta}} = \mathsf{E}\left[\delta\boldsymbol{\vartheta}\delta\boldsymbol{\vartheta}^T\right] = \boldsymbol{C}^{-1}\boldsymbol{B}\boldsymbol{C}^{-T}$, where



(a)



(b)

Figure 5.6: (a) PWVD of the 4^{th} order PPS signal. (b) Simulated and theoretical SNR_{out} curves: $L = 3$; $(-\cdot)$ theory, $(\circ)$ simulation. $L = 21$; $(\cdots)$ theory, $(\diamond)$ simulation. $L = 41$; $(--)$ theory, $(*)$ simulation. Solid line shows asymptotic SNR_{out} bound.

$$\begin{aligned}\boldsymbol{C} &= |A|^2 \sum_{n=M}^{N-M-1} \sum_{l=-M}^{M} \left[-j2l\, \frac{\partial^2 \tilde{\omega}(n;\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}^T}\bigg|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}_0} \right. \\ &\quad \left. +4l^2 \left[\frac{\partial\tilde{\omega}(n;\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}} \frac{\partial\tilde{\omega}(n;\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}^T}\right]\bigg|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}_0}\right], \qquad (5.16)\end{aligned}$$

$$\begin{aligned}\boldsymbol{B} &= 8|A|^2\sigma_V^2 \sum_{n=M}^{N-M-1} \sum_{m=M}^{N-M-1} \left[\left[\frac{\partial\tilde{\omega}(n;\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}} \frac{\partial\tilde{\omega}(m;\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}^T}\right]\bigg|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}_0} \right. \\ &\quad \left. \times \sum_{l=-M}^{M} \sum_{k=-M}^{M} l\, k\, \delta(n-m+l-k)\right] \\ &\quad +\frac{4}{3}\sigma_V^4 M(M+1)(2M+1) \sum_{n=M}^{N-M-1} \left[\frac{\partial\tilde{\omega}(n;\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}} \frac{\partial\tilde{\omega}(n;\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}^T}\right]\bigg|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}_0} \qquad (5.17)\end{aligned}$$

The derivation of the above formulae is included in Appendix 5.B. From (5.16) and (5.17) one can easily show that

$$\boldsymbol{\Gamma}_{\delta\boldsymbol{\vartheta}} = \frac{1}{SNR_{in}}\boldsymbol{D} + \frac{1}{SNR_{in}^2}\boldsymbol{E},$$

where the matrices $\boldsymbol{D}$ and $\boldsymbol{E}$ depend on M, N and the first and second order derivatives of the IF model at $\boldsymbol{\vartheta}_0$. We note that the above expression for $\boldsymbol{\Gamma}_{\delta\boldsymbol{\vartheta}}$ has the same form as the variance of the WHT estimator given in [10], for linear FM signals. Further, for particular IF models such as PPS, it is evident from (5.16) and (5.17) that the variance is independent of the true signal parameter value $\boldsymbol{\vartheta}_0$.

A Monte Carlo simulation was conducted to verify theoretical results for a number of signal models. In Figures 5.7, 5.8 and 5.9 we show the simulated variance and root mean-squared error (RMSE) for linear FM (LFM), quadratic FM (QFM) and sinusoidal FM (SFM) signal models respectively. The phase function for each case is defined as

$$\begin{aligned}\psi_{LFM}(n;\boldsymbol{\vartheta}) &= an + b/2n^2; & \boldsymbol{\vartheta} &= (a,b)^T \\ \psi_{QFM}(n;\boldsymbol{\vartheta}) &= an + b/2n^2 + c/3n^3; & \boldsymbol{\vartheta} &= (a,b,c)^T \\ \psi_{SFM}(n;\boldsymbol{\vartheta}) &= B\sin(\omega_0 n + \phi); & \boldsymbol{\vartheta} &= (B,\omega_0,\phi)^T.\end{aligned}$$

All simulations were conducted with 500 Monte Carlo runs, $N = 128$ samples and PWVD window lengths 27, 21 and 11 respectively. The Cramér-Rao Bound (CRB) for each case is also shown as a reference[4]. We note that the simulated variance coincides with the theoretical expression in all cases. However, the MSE

---

[4] A general form of the CRB for the signal model considered here is derived in Appendix 5.C.

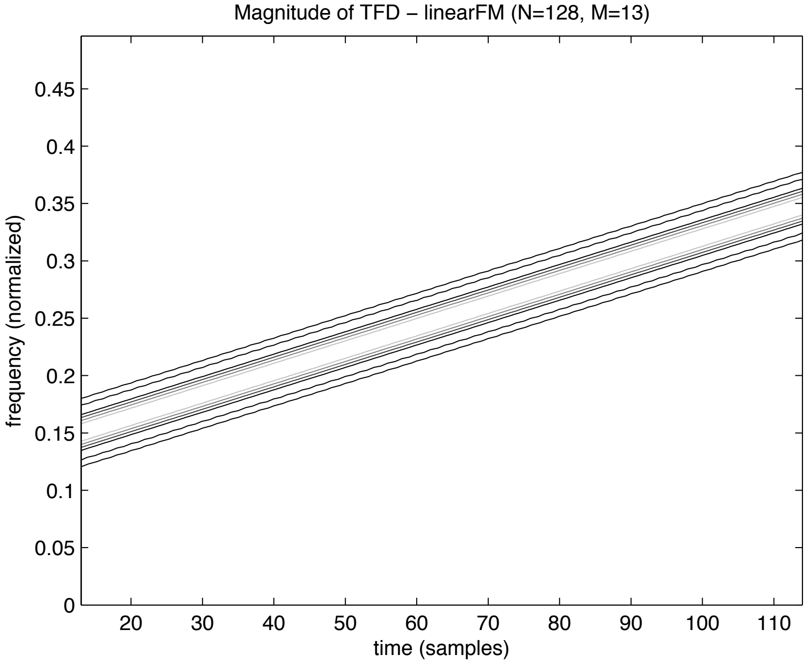
for the amplitude parameter, B, of the sinusoidal FM model and the first-order phase parameter, a, of the quadratic FM model, differ from the variance at higher SNR. This bias results from the nonlinearity of the IF and can be reduced by using a smaller window length, at the expense of increased variance.

For the particular case of linear FM signals, we have also obtained closed form expressions for the estimator variance. The signal parameters of interest are denoted a_0 and b_0, corresponding to the mean frequency and the sweep rate, respectively. The respective estimators based on the PWHT are denoted $\hat{a}$ and $\hat{b}$, where:

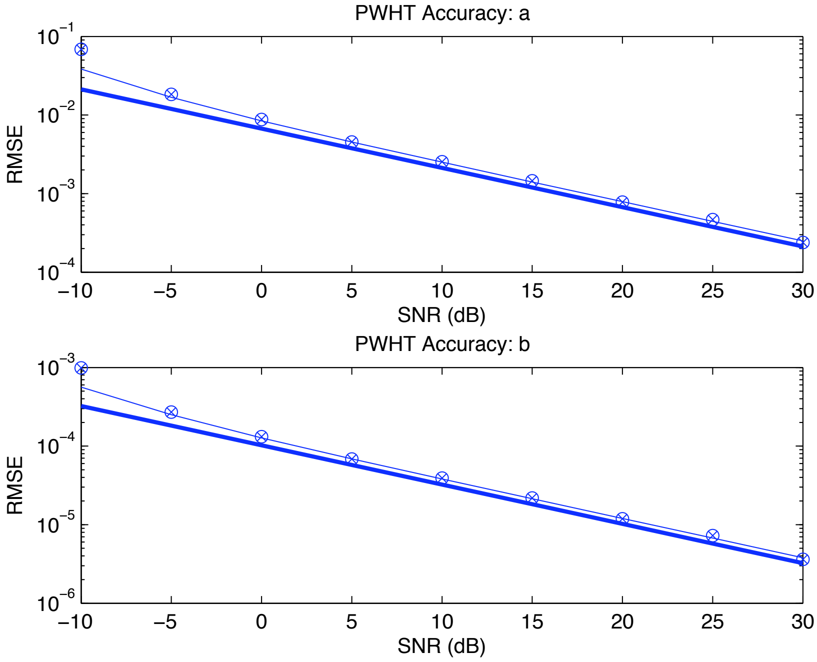
$$\begin{aligned}
\mathsf{Var}\left[\hat{a}\right] \;=\;& \Big\{\frac{12}{SNR_{in}}\big[112(M^6 - M^5(2N-1)) \\
&+(N-1)^2(35N^2+7N-1) \\
&-4M^3(77N^3-177N^2+95N+5) \\
&+7M^2(10N^4-62N^3+91N^2-28N-3) \\
&+M(70N^4-280N^3+283N^2-62N-11) \\
&+8M^4(39N^2-64N+18)\big] \\
&-\frac{105}{SNR_{in}^2}\big[4M^3(13N^2-6N+1) \\
&+2M(7N^4-9N^3+3N-1) \\
&(16M^5-40M^4N-(N-1)^2N(2N^2+N-1) \\
&-2M^2N(19N^2-18N+3)\big]\Big\} \times \Big\{70M(M+1)(2M+1) \\
&\times(N-2M)^3(N^2-4MN+4M^2-1)^2\Big\}^{-1} \qquad (5.18)
\end{aligned}$$

$$\begin{aligned}
\mathsf{Var}\left[\hat{b}\right] \;=\;& \Big\{\frac{108}{SNR_{in}}\big[12(-1+32M^4+M^3(8-56N) \\
&+14M^2(N-4)N+7N^2+2M(-2-14N+7N^2)\big] \\
&-\frac{315}{SNR_{in}^2}\big[8M^3+N-12M^2N-N^3+M(-2+6N^2)\big]\Big\} \\
&\times\Big\{35M(1+3M+2M^2)(8M^3+N-12M^2N-N^3+ \\
&M(-2+6N^2))^2\Big\}^{-1} \qquad (5.19)
\end{aligned}$$

While the above expressions are somewhat complicated, it can be shown that the estimator is not asymptotically efficient. To illustrate this we have plotted the ratio of the variance of $\hat{a}$ and $\hat{b}$ to the CRB, in Figure 5.10. In these plots we have considered only the dominant SNR term when $SNR_{in} > 1$, i.e. ignored the term proportional to $1/SNR_{in}^2$. We note that although the estimation is not efficient, there is clearly an optimal value of M for which the variance is minimized, which occurs around $M = 0.1N$.

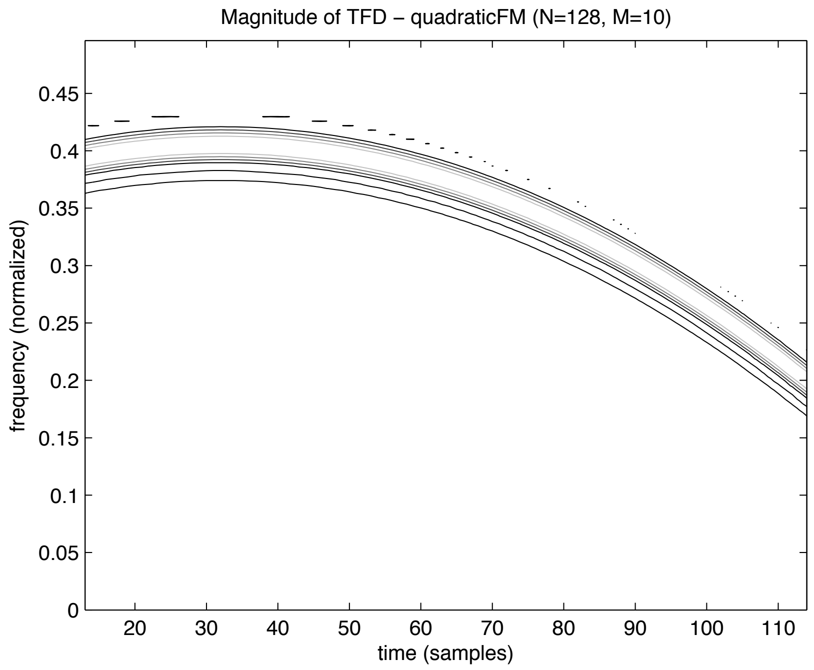


(a)

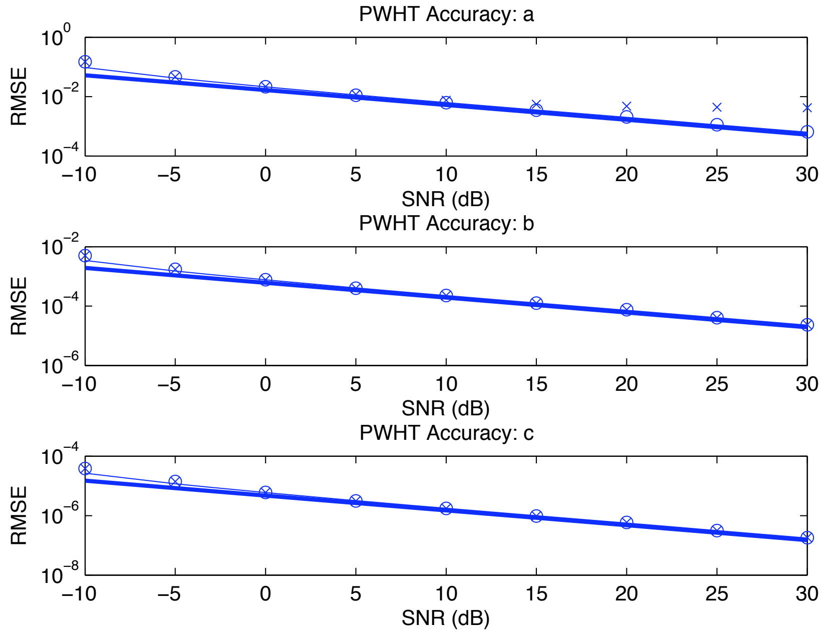


(b)

Figure 5.7: Theoretical and simulated accuracy of the PWHT estimator for a LFM signal model. (a) PWVD of the signal. (b) Accuracy: theoretical variance (-) CRB (**-**) simulated variance (○) simulated RMSE (×).

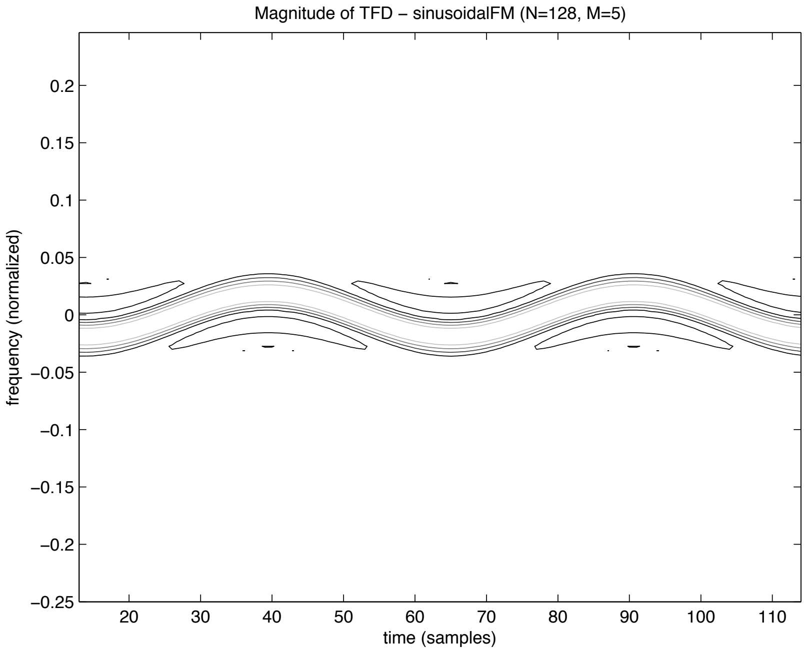


(a)

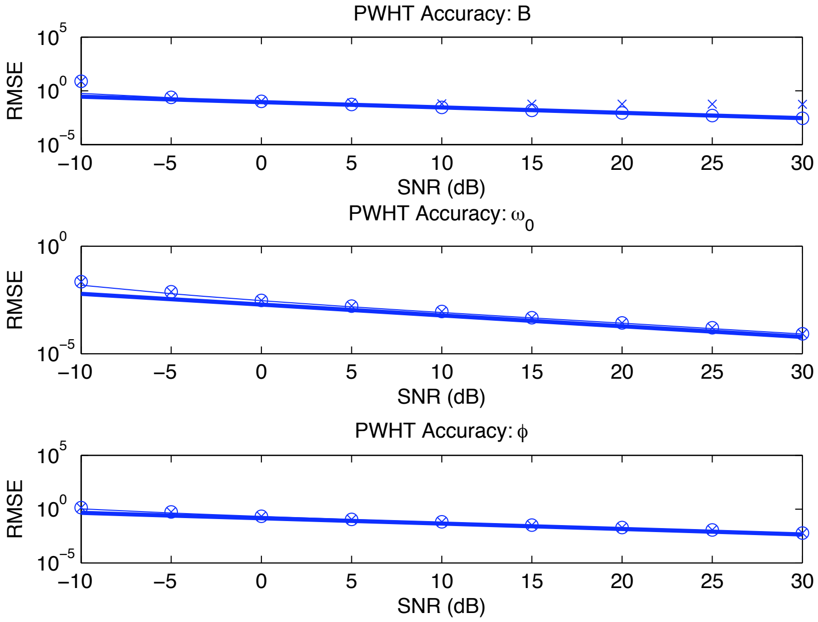


(b)

Figure 5.8: Theoretical and simulated accuracy of the PWHT estimator for a QFM signal model. (a) PWVD of the signal. (b) Accuracy: theoretical variance (-) CRB (**-**) simulated variance (○) simulated RMSE (×).



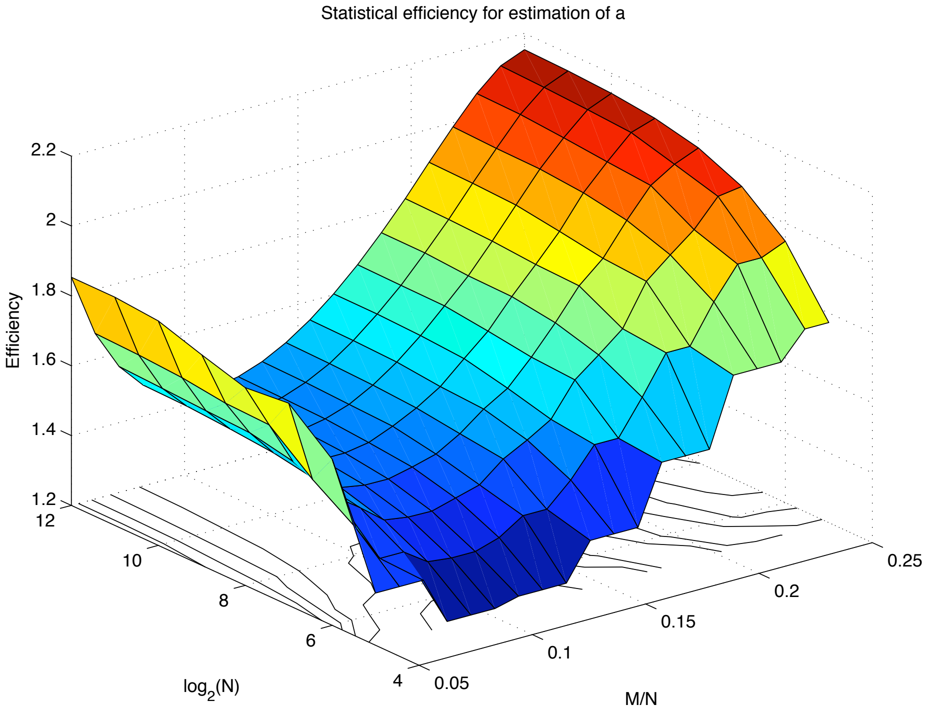
(a)



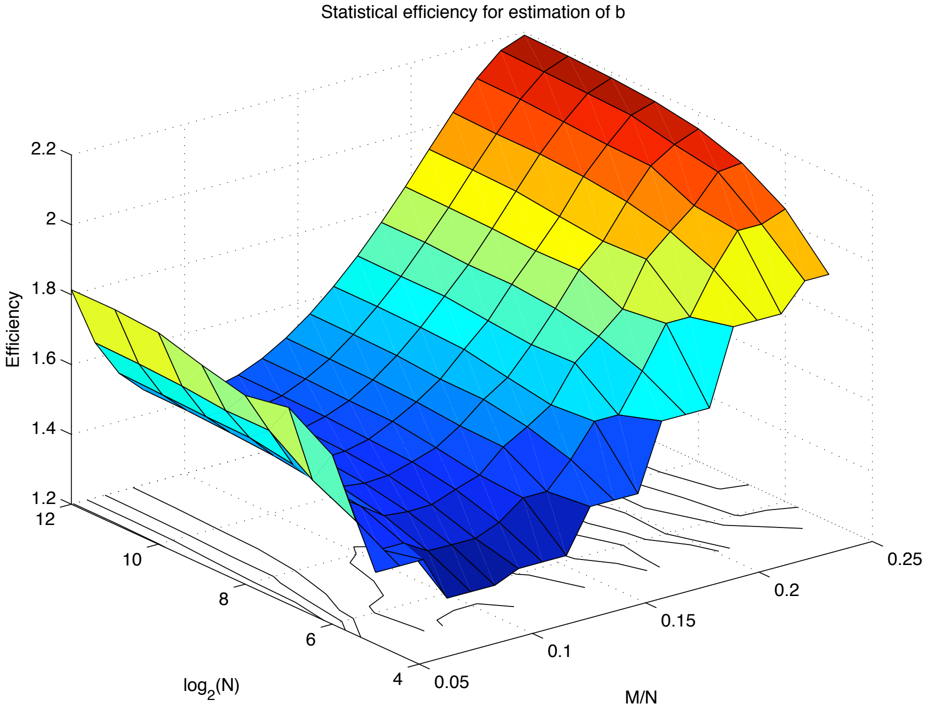
(b)

Figure 5.9: Theoretical and simulated accuracy of the PWHT estimator for a SFM signal model. (a) PWVD of the signal. (b) Accuracy.

Since M is real-valued, the optimal value of M which gives the lowest variance may be obtained by differentiating (5.18) and (5.19) with respect to M and solving



(a)



(b)

Figure 5.10: Efficiency of the PWHT estimator for mean frequency (a) and chirp-rate (b).

for the location of the minimum variance in terms of N. However, a direct algebraic solution of the above equations is somewhat difficult due to the high order terms present. Since the minimum is known to occur close to $M = 0.1N$, we have taken a first order Taylor series expansion of the derivatives of (5.18) and (5.19) about this location, and solved for M. The approximate expression of the optimal value of M for a and b respectively, were found to be

$$\begin{aligned} M_{opt,a} &= -0.311009 + \frac{1}{SNR_{in}} 2.69542 + 0.103864N + O(1/N), \\ M_{opt,b} &= -0.344559 + \frac{1}{SNR_{in}} 2.64928 + 0.106091N + O(1/N). \end{aligned}$$

In many practical cases of interest, e.g. for $SNR_{in} > 1$ and $N >> 27$, the optimal value of M is close to $0.1N$ for estimation of both a and b and one may use this value as a rule of thumb.

5.1.3.4 Implementation

The proposed estimation algorithm is summarised in Table 5.2 and described as follows: Firstly, one chooses a set of window lengths $M_1 < M_2 < \cdots < M_p$ and an initial estimate of the strongest signal component is obtained. This may be done, e.g, via a grid search of the PWHT with $M = M_1$. The appropriate grid spacing in the case of linear FM signals is determined from the RoA expressions given in Section 5.1.3.1. The estimate is then refined in steps via optimization of the PWHT using $M_1, \ldots, M_q$ successively, where the estimate obtained using $M = M_i$ serves to initialize the optimization with $M = M_{i+1}$. In the case of linear FM signals, one may use the final estimate ($M = M_q$) to initialize optimization of the WHT function.

1. Define $M_1 < M_2 < \cdots < M_p$.
2. Perform grid search of using $M = M_1$ to obtain $\hat{\boldsymbol{\vartheta}}_0$. Set $i \leftarrow 1$.
3. Obtain $\hat{\boldsymbol{\vartheta}}_i$, via gradient-based optimization of (5.6), with initial location $\hat{\boldsymbol{\vartheta}}_{i-1}$.
4. Set $i \leftarrow i + 1$. While $i <= p$ repeat from 3.
5. If a linear FM model is applies, use $\hat{\boldsymbol{\vartheta}}_p$ to initialize optimization of the WHT, otherwise $\hat{\boldsymbol{\vartheta}} = \hat{\boldsymbol{\vartheta}}_p$.

Table 5.2: Estimation algorithm based on the PWHT.

For optimization of the PWHT using efficient gradient based schemes, one may require both the first and second order derivatives with respect to the parameters. We first reformulate the PWHT given in Equation (5.6), in a more compact matrix equation:

$$P_{xx}(\boldsymbol{\vartheta}) = \mathbf{1}'(\boldsymbol{R}_{xx} \odot \boldsymbol{W}(\boldsymbol{\vartheta}))\mathbf{1} \tag{5.20}$$

where $\mathbf{1}$ is a length L vector of ones and $\boldsymbol{R}_{xx}$ and $\boldsymbol{W}(\boldsymbol{\vartheta})$ are $L \times N - L + 1$ complex-valued matrices given by

$$\begin{aligned} [\boldsymbol{R}_{xx}]_{pq} &= x(q+p-2)x^*(q-p+L-1) \\ [\boldsymbol{W}(\boldsymbol{\vartheta})]_{pq} &= \exp[-j2\tilde{\omega}(q+M-1;\boldsymbol{\vartheta})(p-M-1)] \end{aligned}$$

for $p = 1, \ldots, L$ and $q = 1, \ldots, N - L + 1$. The first and second order derivatives of $P(\boldsymbol{\vartheta})$ with respect to the parameters are then given by

$$\frac{\partial P(\boldsymbol{\vartheta})}{\partial \vartheta_i} = \mathbf{1}'(\boldsymbol{R}_{xx} \odot \boldsymbol{W}(\boldsymbol{\vartheta}) \odot \boldsymbol{G}_i(\boldsymbol{\vartheta}))\mathbf{1} \tag{5.21}$$

$$\frac{\partial^2 P(\boldsymbol{\vartheta})}{\partial \vartheta_i \partial \vartheta_k} = \mathbf{1}'(\boldsymbol{R}_{xx} \odot (\boldsymbol{G}_i(\boldsymbol{\vartheta}) \odot \boldsymbol{G}_k(\boldsymbol{\vartheta}) + \boldsymbol{W}(\boldsymbol{\vartheta}) \odot \boldsymbol{K}_{ik}(\boldsymbol{\vartheta})))\mathbf{1} \tag{5.22}$$

where

$$\begin{aligned} [\boldsymbol{G}_i(\boldsymbol{\vartheta})]_{pq} &= -j2\frac{\partial \tilde{\omega}(q+M-1;\boldsymbol{\vartheta})}{\partial \vartheta_i}(p-M-1) \\ [\boldsymbol{K}_{ik}(\boldsymbol{\vartheta})]_{pq} &= -j2\frac{\partial^2 \tilde{\omega}(q+M-1;\boldsymbol{\vartheta})}{\partial \vartheta_i \partial \vartheta_k}(p-M-1) \end{aligned}$$

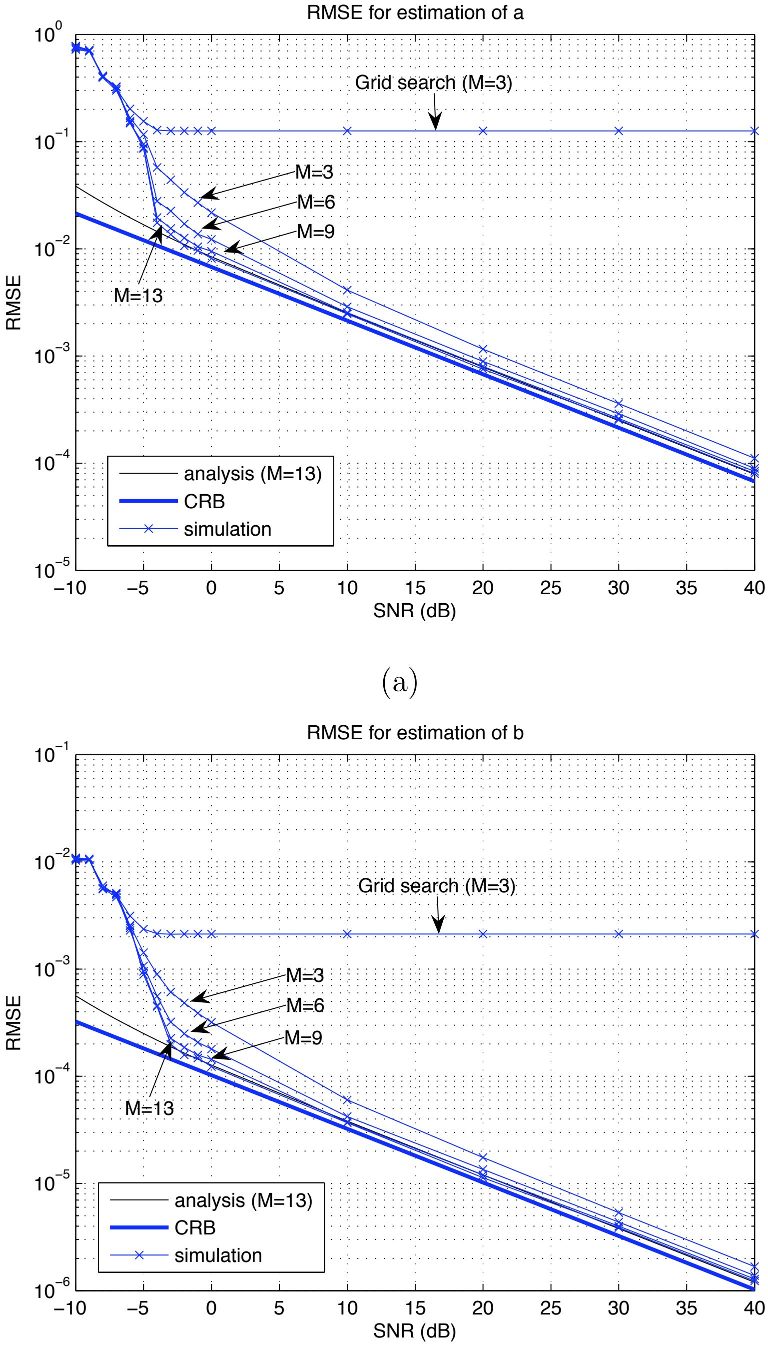
We can see that to compute the gradient vector and Hessian matrix, we require the first and second order derivatives of the IF model. We also notice from Equation (5.21) and Equation (5.22) that the data matrix $\boldsymbol{R}_{xx}$ and the weighting matrix $\boldsymbol{W}(\boldsymbol{\vartheta})$ need only be computed once. From the matrix formulation in Equation (5.20) it is clear that there is a separation between the data and model based calculations. For example, one must only calculate the data matrix $\boldsymbol{R}_{xx}$ once for a given set of observations. A set of trajectories $\{\boldsymbol{\vartheta}_k\}$ through the TF plane corresponding to a rough grid search of the parameter space may be precomputed and the weighting matrices $\{\boldsymbol{W}(\boldsymbol{\vartheta}_k)\}$ stored in memory.

To illustrate the performance of the PWHT estimator with the implementation strategy given in Table 5.2, we take the example of an LFM signal with mean frequency $a_0 = 0.12/T_s$ and chirp rate $b_0 = 0.23/(NT_s)$, where T_s denotes the sample period. The root mean square error (RMSE) of the estimator for a and b is simulated and compared with theoretical variance expressions and the CRB, as shown in Figure 5.11, with $N = 128$. The algorithm of Table 5.2 is applied, with the window parameter M values $(3, 6, 9, 13)$. The initial estimate is computed by a grid search of the PWHT with $M = 3$. The parameter space is sampled in the range $a \in [0, \pi)$ and $b \in [-a/N, (\pi - a)/n]$ with spacing $\Delta_a/2$ and $\Delta_b/2$ along a and b direction respectively, where Δ_a and Δ_b are given by (5.12) and (5.13), respectively. This sampling scheme requires pre-calculation of a total of 50 PWHT trajectories. We note that a grid search using the same sampling scheme for the WHT (using Δ_a and Δ_b as given in Section 5.1.2) would require 4736 trajectories. In optimization of the PWHT we have used an efficient gradient based technique proposed by Fletcher and Powell (FP) [60] for an SNR of 0 dB and above. It was found that for an SNR below 0 dB, the FP algorithm did not always converge and in these cases we have used a more robust, albeit more complex, algorithm proposed by Nelder and Mead (NM) [88], which was found to produce good results down to about -5 dB SNR. The simulation results have been obtained by averaging 500 Monte Carlo runs.

In order to illustrate the effect of increasing the window length, in the procedure of Table 5.2, the RMSE for each value of M is plotted in Figure 5.11. While the initial accuracy of the grid search (with $M = 3$) is very poor, it is sufficient for initialization of the optimization routine, which provides increasing accuracy as the value of M is increased. We note that the simulated accuracy shown in Figure 5.11 is consistent with theoretical analysis (for $M = 13$), down to about -3 dB SNR. The discrepancy at extremely low SNR is expected as assumptions inherent in the derivation of the RoA are no longer valid. As the noise floor increases, the true RoA shrinks, and for SNR << 1 one needs to use a denser grid search in order to achieve convergence. For higher SNR, although the estimation is not efficient, the performance is still very close to the CRB and the computation time has been greatly reduced when compared to the WHT based estimator. In this case, the need to use the WHT becomes questionable, since the large increase in computational burden provides only a minor improvement in estimation accuracy. However, statistical efficiency is easily achieved, if desired, by optimising the WHT using the PWHT estimate for initialising the search. The overall approach is still far more computationally efficient than trying to directly optimise the WHT function.

5.1.4 Multicomponent Signal Estimation

In the case of multicomponent signals, there will be a number of peaks within the parameter space of the PWHT, as illustrated in Figure 5.3. One approach in this case, is to threshold the (P)WHT function to determine the number and/or



(a)

(b)

Figure 5.11: Single-component linear FM estimation performance; simulated, theoretical and the CRB. (a) RMSE for a. (b) RMSE for b.

rough location of the peaks. However, such a method would require a relatively fine grid search despite the increased peak width of the PWHT, and the issue of how to set an appropriate threshold must be considered. From a statistical point of view, one may formulate the peak search as a detection problem, and set the threshold based on a certain level of significance. This then requires knowledge of the probability distribution of the (P)WHT.

To avoid these issues, we propose sequential estimation of each component. One estimates the ‘strongest’ component from the largest peak of the PWHT, using the method outlined in Table 5.2. The complex amplitude is then estimated, for example using a simple least-squares approach, and the reconstructed component is then subtracted from the observations. This is repeated until all components have been estimated. If the number of components is unknown, one may construct a test to determine when the residual term contains no more signal components, though this is not elaborated upon here. The estimation procedure is summarised in Table 5.3.

1. Set $i \leftarrow 1$. Organize the observations into the vector $\boldsymbol{x} = (x(0), \ldots, x(N-1))'$.
2. Estimate $\boldsymbol{\vartheta}_i$ using the procedure of Table 5.2, to obtain $\hat{\boldsymbol{\vartheta}}_i$.
3. Form the vector $\boldsymbol{s}_i = (e^{j\psi(0;\hat{\boldsymbol{\vartheta}}_i)}, \ldots, e^{j\psi(0;\hat{\boldsymbol{\vartheta}}_i)})'$.
4. Estimate the amplitude according to $\hat{A}_i = \boldsymbol{s}^H \boldsymbol{x}/N$.
5. Remove the estimated component: $\boldsymbol{x} \leftarrow \boldsymbol{x} - \hat{A}_i \boldsymbol{s}_i$.
6. Set $i \leftarrow i+1$. If $i \leq d$, repeat from Step 2.

Table 5.3: Multicomponent estimation algorithm based on the PWHT.

In contrast to sequential phase-based estimation of multicomponent signals [89], we do not require that the component being estimated is stronger than the other components present. However, the proposed approach suffers from an inherent bias since the multiple peaks disturb one another through main and/or side lobes, resulting in a shift of the peak location from the true parameter values, even if Assumption 5 holds. In order to reduce this bias, one may obtain the initial estimates $\{\hat{\boldsymbol{\vartheta}}_i\}_{i=1}^{q}$ using the procedure of Table 5.3, followed by a bias reduction step, in which all estimated components but the desired are subtracted from the observations, and the desired component parameters are then re-estimated using the PWHT. In this bias reduction step, one may use the previous (biased) estimate to initialize optimization of the PWHT with the final value $M = M_q$, i.e. it is not necessary to repeat the full estimation procedure of Table 5.2 again.

To demonstrate the multicomponent estimation procedure, we first consider a multicomponent LFM signal, of which the PWVD and PWHT is shown in Figure 5.12. Although the PWVD contains significant cross-terms, the PWHT contains only three clear peaks. In the implementation of the procedure from Table 5.2 (in step 2 of Table 5.3), we have used the same grid-search points as in the previous section. We have also applied the WHT after the last PWHT step, with two final iterations to reduce bias. Figure 5.13 shows the simulated estimation accuracy using $N = 128$ observations. It can be seen from the results that after two bias reduction iterations, the estimation accuracy becomes close to that of the single component case. Of course, at extremely high SNR when the bias becomes the dominant source of estimation error, more bias reduction iterations are typically required to achieve statistical efficiency.

5.2 Application to DOA estimation

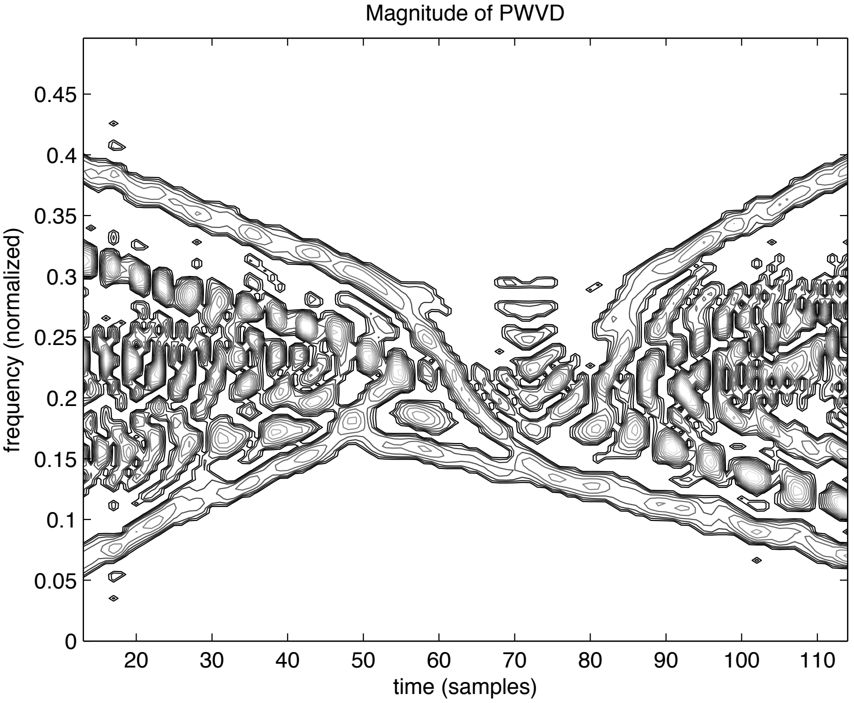
Let us first recall the idea by Amin *et al* [7] for direction finding based on a spatial TFD matrix, which was reviewed in Section 3.2.3. The estimate of the DOAs, $\boldsymbol{\theta}$, for a subset of $d_0 \leq d$ signals is obtained via subspace decomposition of an averaged STFD matrix. The averaged STFD matrix is defined as:

$$\boldsymbol{D} = \frac{1}{d_o N_o} \sum_{q \in \mathcal{S}_{d_o}} \sum_{n \in \mathcal{S}_{N_o}} \boldsymbol{D}_{\boldsymbol{XX}}(n, \tilde{\omega}_q(n)). \tag{5.23}$$

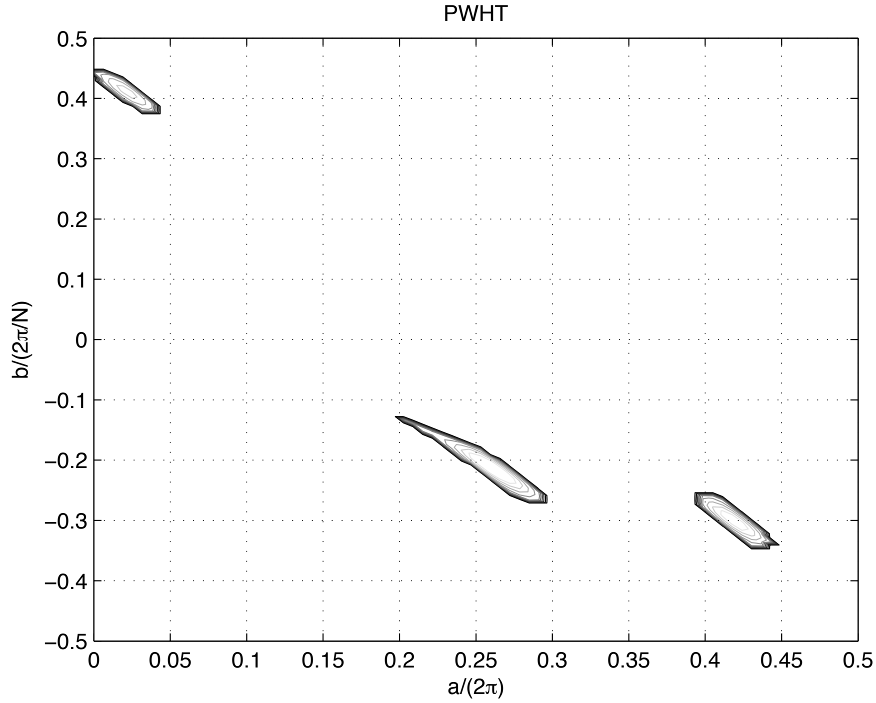
where $\{\tilde{\omega}_q(n)\}_{q=1}^{d}$ are the IF functions of the source signals. Compared with the sample covariance matrix of the array output, the matrix in Equation (5.23) provides an effective improvement in SNR by amplification of the source Eigen-values with respect to the noise Eigen-values [124]. We note that direction finding based on Equation (5.23) is performed under the constraint $d_0 < m$, where m is the number of sensors. The best estimator performance is achieved by using $d_0 = 1$, and successively estimating the DOAs of each source.

5.2.1 Direction Finding with the Hough transform

The use of Equation (5.23) requires knowledge of the source signal IFs $\{\tilde{\omega}_k(t)\}_{k=1}^{d}$ for each corresponding direction parameter to be estimated [124]. In order to implement such a scheme without *a priori* knowledge of the signal IFs, we may apply the TF signautre estimation method described earlier in this chapter. This approach requires us to know a general functional form that is suitable for describing the IF, but not the particular parameters for the signals being observed. By



(a)



(b)

Figure 5.12: Multicomponent linear FM signal example: (a) Threshold-ed PWVD of the noise-free signal. (b) Threshold-ed PWHT of the noise-free signal.

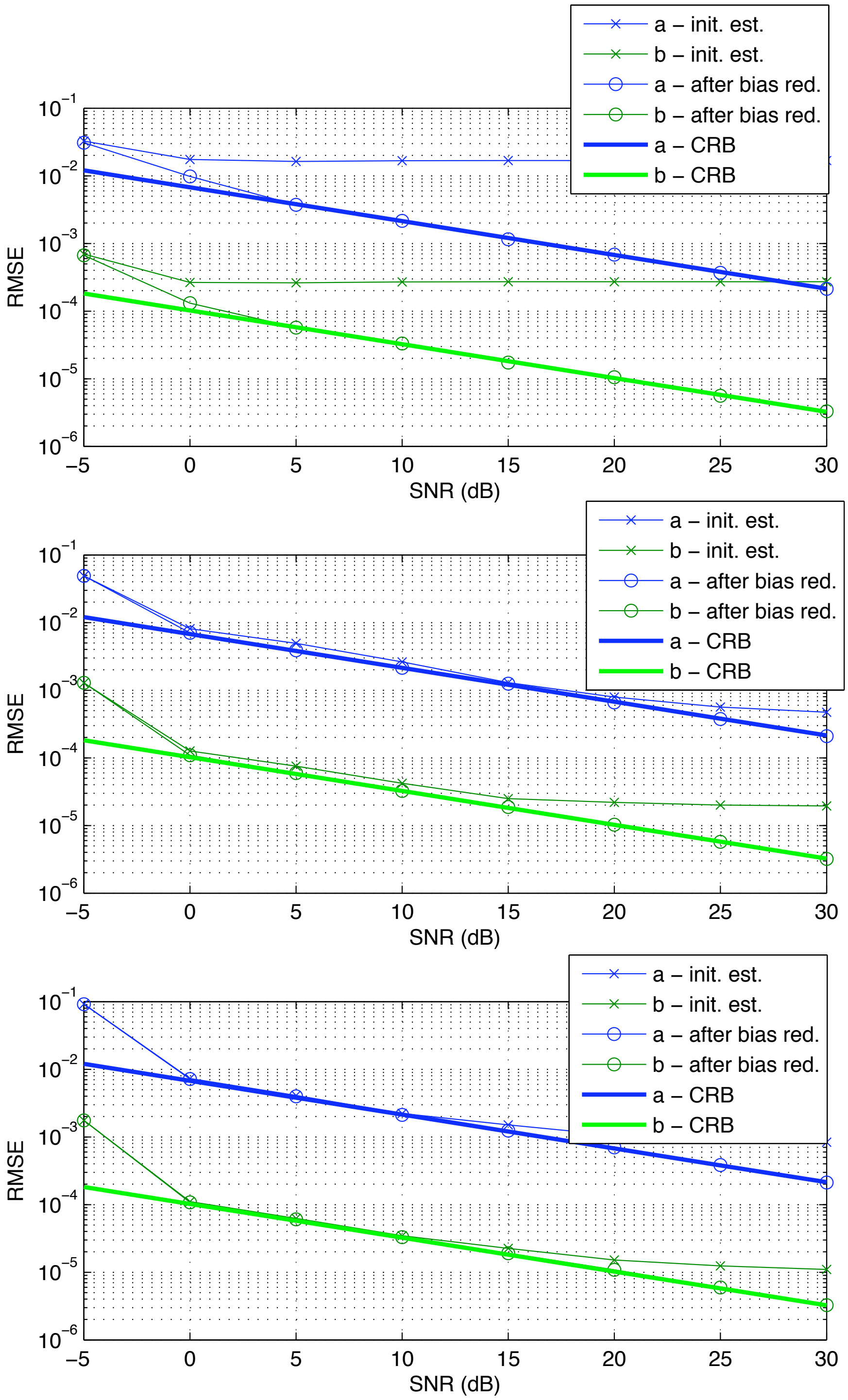


Figure 5.13: Multicomponent linear FM signal example. RMSE of the estimator; simulated, theoretical and the CRB. From top to bottom corresponds to the first, second and third estimated components respectively.

parameterizing the signal IFs we can translate the problem of estimating $\{\omega_k(t)\}$ into peak detection in a parameter space.

Denoting the parametrized IFs by $\{\tilde{\omega}(t;\boldsymbol{\vartheta}_k)\}_{k=1}^{d}$, where $\boldsymbol{\vartheta}_k$ is the vector of IF parameters for source k and choosing $d_0 = 1$ in Equation (5.23) for best performance, we obtain

$$\boldsymbol{H}_{\boldsymbol{XX}}(\boldsymbol{\vartheta}_k) = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{D}_{\boldsymbol{XX}}(n, \tilde{\omega}(n;\boldsymbol{\vartheta}_k)), \quad k = 1, \ldots, d \tag{5.24}$$

which we call a spatial time-frequency Hough transform (STFHT) matrix. In order to use the Hough transform for estimating $\{\boldsymbol{\vartheta}_k\}_{k=1}^{d}$, we first average the auto-sensor TFDs across the array sensor data. This is known to reduce the effect of noise and cross-source components in the TF-plane [87]. The Hough transform of the array averaged TFD is given by

$$\begin{aligned} H_{XX}(\boldsymbol{\vartheta}) &= \frac{1}{mN_o}\sum_{n\in\mathcal{S}_{N_o}}\sum_{k=1}^{m} D_{X_kX_k}(n, \tilde{\omega}(n;\boldsymbol{\vartheta})) \\ &= \frac{1}{m}\sum_{k=1}^{m} H_{X_kX_k}(\boldsymbol{\vartheta}), \end{aligned} \tag{5.25}$$

where $H_{X_kX_k}(\boldsymbol{\vartheta})$ denotes the time-frequency Hough transform of the waveform at sensor k. The estimates $\{\hat{\boldsymbol{\vartheta}}_k\}_{k=1}^{d}$ are obtained from Equation (5.25) as the values of $\boldsymbol{\vartheta}$ corresponding to the d largest peaks in $H_{\boldsymbol{XX}}(\boldsymbol{\vartheta})$. In the case of an unknown number of sources, it would be necessary to appropriately threshold $H_{\boldsymbol{XX}}(\boldsymbol{\vartheta})$ before applying a peak-search algorithm. However the details and performance of such a source number detection scheme will not be discussed here.

Direction of arrival estimation for the signal $s_k(t)$ is conducted by first obtaining estimates $\hat{\boldsymbol{\vartheta}}_k$ from Equation (5.25) and forming the corresponding STFHT matrix according to Equation (5.24). A high resolution second-order direction finding method such as MUSIC can be applied. We note that obtaining $\{\hat{\theta}_k\}_{k=1}^{d}$ requires searching for a single peak d times as opposed to searching for d peaks in, e.g. a conventional MUSIC spectrum. By appropriate choice of the IF model, we also have the possibility of only estimating $\{\hat{\theta}_k\}$ for those sources with, e.g. a strongly linear IF such as chirp signals and ignoring others with non-localized TF representations. If the IFs of the different source signals do not significantly overlap, then each peak in the Hough transform isolates the energy of a single signal. The use of the matrices in Equation (5.24) then potentially allows direction finding for an arbitrary number of source signals, using only two sensors. The DOA estimation procedure is summarized in Table 5.4.

1. Determine an appropriate parametrization ($\boldsymbol{\vartheta}$) for the signal IF.
2. Compute $H_{\boldsymbol{XX}}(\boldsymbol{\vartheta})$, the array averaged TFHT, according to Equation (5.25).
3. Form the estimates $\{\hat{\boldsymbol{\vartheta}}_k\}_{k=1}^d$ from the values of $\boldsymbol{\vartheta}$ corresponding to the d largest peaks of $H_{\boldsymbol{XX}}(\boldsymbol{\vartheta})$.
4. Calculate the STFHT matrix $\boldsymbol{H}_{\boldsymbol{XX}}(\hat{\boldsymbol{\vartheta}}_k)$ for $k = 1, \ldots, d$.
5. Form the DOA estimates $\{\hat{\theta}_k\}_{k=1}^d$ by applying MUSIC or other subspace based technique to the matrices formed in Step 4.

Table 5.4: DOA estimation algorithm using the Hough transform.

5.2.2 Simulations

In the following simulations, we focus on the chirp signal, whose Hough transform corresponds to taking line integrals through the TF plane. The signal IFs are given by $\{\tilde{\omega}(t, \boldsymbol{\vartheta}_k) = 2\pi(a_k + b_k t)\}_{k=1}^d$ where the parameter vector in the TFHT is $\boldsymbol{\vartheta}_k = [a_k, b_k]$. The SNR for signal k is defined as A_k^2/σ^2.

To illustrate the effect of the array averaging, we show in Figure 5.14 the Hough transform at a reference sensor and that averaged across the array. In this example two chirp signals are impinging on an eight-sensor ULA, each with an SNR of -20 dB and respective DOAs of -10 and 10 degrees. 1024 snapshots are used and the TFD is a pseudo Wigner-Ville distribution with odd window length of 129 samples. Clearly the noise floor over the parameter space is reduced due to array averaging, and the two peaks due to the two chirp signals are enhanced. We note also that due to their oscillatory nature, the cross-terms are significantly reduced by the integration through the TF plane when calculating the TFHT. This is in addition to the reduction of cross-source components achieved by the array averaging.

In the next example, we show the overall performance in DOA estimation, for the signal mixture defined previously. Figure 5.15 shows the root mean-squared error (RMSE) obtained from averaging 200 Monte Carlo runs, for estimation of the first source's DOA, as obtained by the algorithm in Table 5.4. The estimator performance is simulated vs SNR and compared with the case of known signal

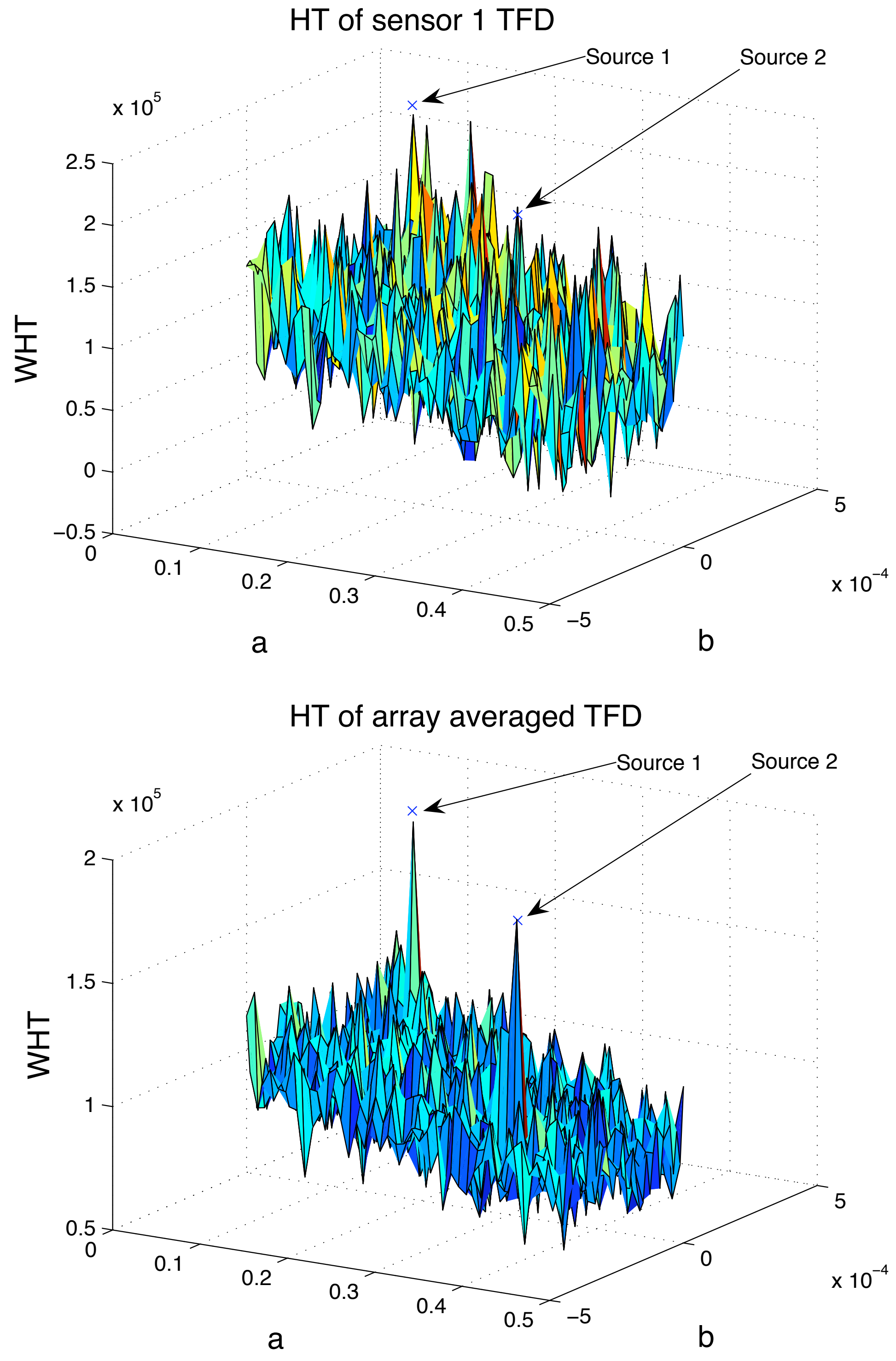


Figure 5.14: Example of TFHT before (top) and after (bottom) the array averaging.

IF, the conventional root-MUSIC algorithm and the CRB. We observe that the proposed algorithm performs as well as the case of known signal IF, for SNR greater than -20 dB. This shows that estimation of the signal IF is sufficiently accurate at low SNR where TF-MUSIC exhibits a significant performance gain over conventional MUSIC.

In the third simulation example, we consider the case of fewer sensors than sources. The same signals previously used are present, plus a third chirp signal, all with SNR of -5 dB. The signal DOAs are given by -12, 2 and 15 degrees respectively and the ULA has only two sensors. Due to the isolation of individual signal TF signatures by the use of the Hough transform, we are able to estimate the DOA of each source individually despite the fact that the system is underdetermined ($d > m$). MUSIC spectra are plotted in Figure 5.16 which illustrate that the three chirp signals are able to be resolved.

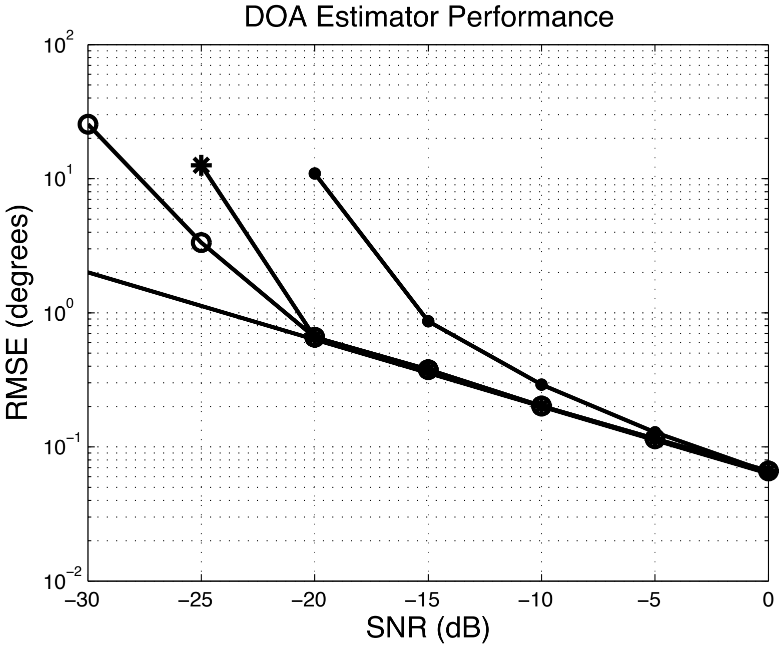


Figure 5.15: RMSE of DOA estimator vs SNR. (-) CRB. (o) TF-MUSIC with known signal IF. (*) TF-MUSIC with estimated signal IF. (·) root-MUSIC.

5.3 Conclusions

With the goal of TF signature estimation in mind, the first part of this chapter examined a computationally attractive implementation of the WHT estimator for linear FM signals. A technique for reducing the number of required trajectories to be computed was proposed based on the PWVD, having window length L, which widens the peak of interest in the parameter space. It was shown that the

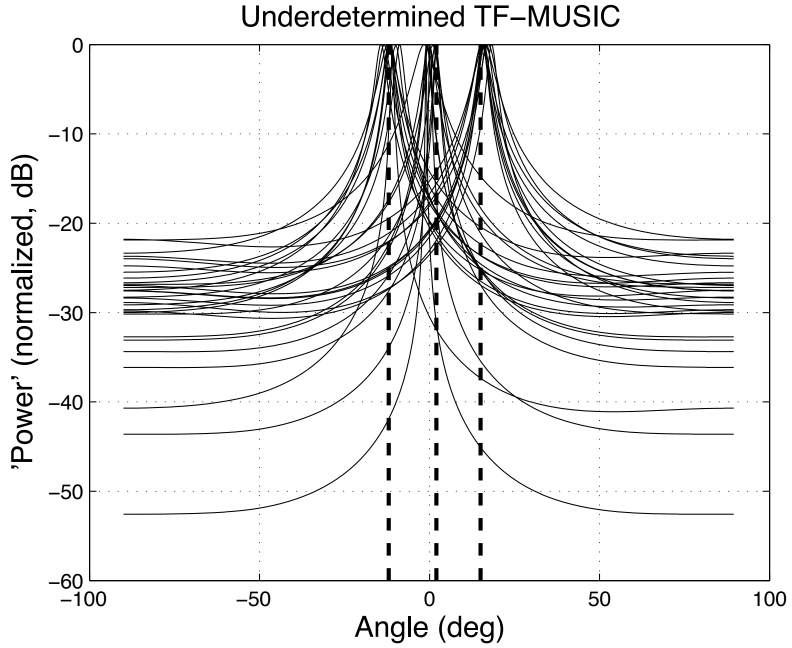


Figure 5.16: MUSIC spectra in the under-determined case ($d = 3 > m = 2$) with signal DOAs $[-12, 2, 15]$ degrees.

peak width is increased from $O(1/N)$ for the WHT to $O(1/L)$ for the PWHT, allowing one to effectively control the RoA of the estimator.

An estimation scheme for both single and multicomponent signals was proposed based on this result. The application of the PWHT to nonlinear FM signals was also considered, and shown to be effective under the assumption that the signal IF is approximately linear within all intervals of length L. Statistical analysis was performed in the case of additive white Gaussian noise, which showed how the SNR performance threshold is affected, and interestingly, that the output SNR is the same as that of the WHT, despite the reduced computational requirement. A general expression for the estimator variance was provided for nonlinear FM signals, and in closed form for linear FM signals.

It was shown that in the case of linear FM signals there is clearly an optimal choice for the PWVD window length which yields a performance level close to that of the CRB. Theoretical results presented were verified with simulations for linear and sinusoidal FM signals, in both the single and multicomponent case.

The second part of this chapter looked at the application of the proposed TF signature estimation to the problem of DOA estimaion. The proposed approach was based on STFD matrices. Prior knowledge of the source TF signatures was not assumed, but estimated according to a parametric model using the TF Hough transform. Two important conclusions of the latter work are as follows; firstly, the use of the Hough transform provides performance close to the case of exactly known TF signatures at low SNR, where there is a significant gain in performance to be achieved by using STFD direction finding, over traditional methods. Secondly, the Hough transform allows DOA estimation for more sources than sensors, in an automatic way.

5.A Derivation of the output SNR

In order to determine the output SNR as defined in Equation (5.14) we need to find $|P_s(\boldsymbol{\vartheta}_0)|^2$ and $\mathsf{Var}\left[P_X(\boldsymbol{\vartheta}_0)\right]$ where $s(n) = Ae^{j\psi(n;\boldsymbol{\vartheta}_0)}$ and $X(n) = s(n) + V(n)$. In the following we assume $V(n)$ to be a complex circular white Gaussian process of variance σ_V^2.

Let us express the PWHT of the observations as $P_X(\boldsymbol{\vartheta}) = P_s(\boldsymbol{\vartheta}) + P_{sV}(\boldsymbol{\vartheta}) + P_V(\boldsymbol{\vartheta})$ where

$$P_{sV}(\boldsymbol{\vartheta}) \triangleq \sum_{n=M}^{N-M-1} \sum_{l=-M}^{M} [s(n+l)V^*(n-l) + s^*(n-l)V(n+l)]e^{-j2\tilde{\omega}(n,\boldsymbol{\vartheta})l} \quad (5.26)$$

and the expressions for $P_s(\boldsymbol{\vartheta})$ and $P_V(\vartheta)$ follow the definition given in Equation (5.6). Since the third order moments of $V(n)$ are zero, then $\mathsf{Var}\left[P_X(\boldsymbol{\vartheta}_0)\right] = \mathsf{Var}\left[P_{sV}(\boldsymbol{\vartheta}_0)\right] + \mathsf{Var}\left[P_V(\boldsymbol{\vartheta}_0)\right]$.

From Equation (5.26) it is clear that $\mathsf{E}\left[P_{sV}(\boldsymbol{\vartheta})\right] = 0$. The variance is calculated according to:

$$\begin{aligned}
\mathsf{Var}\left[P_{sV}(\boldsymbol{\vartheta}_0)\right] &= \mathsf{E}\left[|P_{sV}(\boldsymbol{\vartheta}_0)|^2\right] = \sum_{n=M}^{N-M-1}\sum_{m=M}^{N-M-1}\sum_{l=-M}^{M}\sum_{k=-M}^{M} e^{-j2(\tilde{\omega}(n,\boldsymbol{\vartheta}_0)l - \tilde{\omega}(m,\boldsymbol{\vartheta}_0)k)} \\
&\quad \times \mathsf{E}\big[[s(n+l)V^*(n-l) + s^*(n-l)V(n+l)] \\
&\quad \times [s^*(m+k)V(m-k) + s(m-k)V^*(m+k)]\big] \\
&= \sum_{n=M}^{N-M-1}\sum_{m=M}^{N-M-1}\sum_{l=-M}^{M}\sum_{k=-M}^{M} e^{-j2(\tilde{\omega}(n,\boldsymbol{\vartheta}_0)l - \tilde{\omega}(m,\boldsymbol{\vartheta}_0)k)} \\
&\quad \times [s(n+l)s^*(m+k)\mathsf{E}\left[V^*(n-l)V(m-k)\right] \\
&\quad + s^*(n-l)s(m-k)\mathsf{E}\left[V(n+l)V^*(m+k)\right]].
\end{aligned}$$

Using the circular property of the noise process and Assumption 5, we obtain

$$\begin{aligned}
\mathsf{Var}\left[P_{sV}(\boldsymbol{\vartheta}_0)\right] &\approx |A|^2\sigma_V^2 \sum_{n=M}^{N-M-1}\sum_{m=M}^{N-M-1}\sum_{l=-M}^{M}\sum_{k=-M}^{M} e^{-j2(a_n+b_n n)l} e^{j2(a_m+b_m m)k} \\
&\quad \times [e^{j(a_n n + b_n(n+1)^2/2)}\delta(n-m+k-l) \\
&\quad + e^{j(a_m m + b_m(m+1)^2/2)}\delta(n-m+l-k)],
\end{aligned}$$

where (a_n, b_n) and (a_m, b_m) denote the parameters of the linear IF approximation on the intervals $\mathcal{I}_n$ and $\mathcal{I}_m$ respectively. Since the Kronecker delta functions restrict non-zero terms of the summation to the case $|n-m| < L$, we assume

$a_n \approx a_m$ and $b_n \approx b_m$ in the above equation, allowing the simplification:

$$\begin{aligned}
\mathsf{Var}\left[P_{sV}(\boldsymbol{\vartheta}_0)\right] &\approx |A|^2\sigma_V^2 \sum_{n=M}^{N-M-1}\sum_{m=M}^{N-M-1}\sum_{l=-M}^{M}\sum_{k=-M}^{M} \\
&\quad \left[e^{j(n-m+k-l)(2a_n+b_n(n+m-l-k))/2}\delta(n-m-l+k)+\right. \\
&\quad \left. e^{-j(n-m+l-k)(2a_n+b_n(n+m+l+k))/2}\delta(n-m+l-k)\right] \\
&= |A|^2\sigma_V^2 \sum_{n=M}^{N-M-1}\sum_{m=M}^{N-M-1}\sum_{l=-M}^{M}\sum_{k=-M}^{M} \\
&\quad \left[\delta(n-m-l+k)+\delta(n-m+l-k)\right] \\
&= |A|^2\sigma_V^2\frac{2}{3}(2M+1)(3N(2M+1)-2M(5+8M)) \\
&= |A|^2\sigma_V^2\frac{2}{3}L(3NL-(L-1)(4L+1)).
\end{aligned}$$

The closed form expression in second-to-last line above was obtained with the aid of the software package Mathematica, and is valid under the condition that $N > 4M+1$.

The second order moment of $P_V(\boldsymbol{\vartheta})$ is given by

$$\begin{aligned}
\mathsf{E}\left[|P_v(\boldsymbol{\vartheta})|^2\right] &= \sum_{n=M}^{N-M-1}\sum_{m=M}^{N-M-1}\sum_{l=-M}^{M}\sum_{k=-M}^{M} e^{-j2(\tilde{\omega}(n,\boldsymbol{\vartheta}_0)l-\tilde{\omega}(m,\boldsymbol{\vartheta}_0)k)} \\
&\quad \times\mathsf{E}\left[V(n+l)V^*(n-l)V^*(m+k)V(m-k)\right] \\
&= \sum_{n=M}^{N-M-1}\sum_{m=M}^{N-M-1}\sum_{l=-M}^{M}\sum_{k=-M}^{M} e^{-j2(\tilde{\omega}(n,\boldsymbol{\vartheta}_0)l-\tilde{\omega}(m,\boldsymbol{\vartheta}_0)k)} \\
&\quad \times\left[\mathsf{E}\left[V(n+l)V^*(n-l)\right]\mathsf{E}\left[V^*(m+k)V(m-k)\right]\right. \\
&\quad \left.+\,\mathsf{E}\left[V(n+l)V^*(m+k)\right]\mathsf{E}\left[V^*(n-l)V(m-k)\right]\right]
\end{aligned}$$

Using the circular property of the noise process and Assumption 5, we obtain

$$\begin{aligned}
\mathsf{E}\left[|P_v(\boldsymbol{\vartheta})|^2\right] &\approx \sigma_V^4 \sum_{n=M}^{N-M-1}\sum_{m=M}^{N-M-1}\sum_{l=-M}^{M}\sum_{k=-M}^{M} e^{-j2[(a_n+b_n n)l-(a_m+b_m m)k]} \\
&\quad \times[\delta(l)\delta(k)+\delta(l-k)\delta(n-m)],
\end{aligned}$$

where (a_n, b_n) and (a_m, b_m) denote the parameters of the linear IF approximation on the intervals $\mathcal{I}_n$ and $\mathcal{I}_m$ respectively. Given the arguments of the Kronecker delta functions the above expression simplifies to:

$$\begin{aligned}
\mathsf{E}\left[|P_v(\boldsymbol{\vartheta})|^2\right] &= \sigma_V^4 \sum_{n=M}^{N-M-1}\sum_{m=M}^{N-M-1}\sum_{l=-M}^{M}\sum_{k=-M}^{M}[\delta(l)\delta(k)+\delta(l-k)\delta(n-m)] \\
&= \sigma_V^4((N-2M)^2+(N-2M)(2M+1)) \\
&= \sigma_V^4((N-L+1)^2+(N-L+1)L).
\end{aligned}$$

Also, $\mathsf{E}\left[P_V(\boldsymbol{\vartheta})\right] = \sigma_V^2(N-L+1)$, which leads the overall variance expression

$$\mathsf{Var}\left[P_X(\boldsymbol{\vartheta}_0)\right] = |A|^2\sigma_V^2\frac{2}{3}L(3NL-(L-1)(4L+1)) + \sigma_V^4(N-L+1)L. \quad (5.27)$$

It is easily verified by direction substitution that, under Assumption 5,

$$P_s(\boldsymbol{\vartheta}_0) = |A|^2L(N-L-1). \quad (5.28)$$

Substitution of Equation (5.28) and (5.27) into Equation (5.14) yields the desired expression for the output SNR given in Equation (5.15). ■

5.B Derivation of the estimator bias and variance

The maximum of $P_X(\boldsymbol{\vartheta}) = P_s(\boldsymbol{\vartheta}) + \delta P(\boldsymbol{\vartheta})$ occurs at a location $\boldsymbol{\vartheta}_0 + \delta\boldsymbol{\vartheta}$, where $\boldsymbol{\vartheta}_0$ is the true parameter value and $\delta\boldsymbol{\vartheta}$ is a perturbation due to noise. The maximum of $P_X(\boldsymbol{\vartheta})$ satisfies

$$\left[\frac{\partial}{\partial\boldsymbol{\vartheta}}(P_X(\boldsymbol{\vartheta}) + \delta P(\boldsymbol{\vartheta}))\right]_{\boldsymbol{\vartheta}_0+\delta\boldsymbol{\vartheta}} = \mathbf{0} \quad (5.29)$$

We expand Equation (5.29) using a first order approximation:

$$\left[\frac{\partial}{\partial\boldsymbol{\vartheta}}(P_s(\boldsymbol{\vartheta}) + \delta P(\boldsymbol{\vartheta}))\right]_{\boldsymbol{\vartheta}_0} + \left[\frac{\partial^2}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}^T}P_s(\boldsymbol{\vartheta})\right]_{\boldsymbol{\vartheta}_0}\delta\boldsymbol{\vartheta} \cong \mathbf{0}. \quad (5.30)$$

Under Assumption 5, $[\frac{\partial}{\partial\boldsymbol{\vartheta}}P_s(\boldsymbol{\vartheta})]_{\boldsymbol{\vartheta}_o} = 0$, therefore

$$\delta\boldsymbol{\vartheta} = -\left[\frac{\partial^2}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}^T}P_s(\boldsymbol{\vartheta})\right]_{\boldsymbol{\vartheta}_0}^{-1}\left[\frac{\partial}{\partial\boldsymbol{\vartheta}}\delta P(\boldsymbol{\vartheta}))\right]_{\boldsymbol{\vartheta}_0}. \quad (5.31)$$

Defining $\boldsymbol{C} \triangleq [\frac{\partial^2}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}^T}P_s(\boldsymbol{\vartheta})]_{\boldsymbol{\vartheta}_0}$ and $\boldsymbol{b} \triangleq \left[\frac{\partial}{\partial\boldsymbol{\vartheta}}\delta P(\boldsymbol{\vartheta}))\right]_{\boldsymbol{\vartheta}_0}$, where $\boldsymbol{b}$ is a random vector, the mean and covariance matrix of $\delta\boldsymbol{\vartheta}$ are given by $\mu_{\delta\boldsymbol{\vartheta}} = -\boldsymbol{C}^{-1}\mathsf{E}\left[\boldsymbol{b}\right]$ and $\boldsymbol{\Gamma}_{\delta\boldsymbol{\vartheta}} = \boldsymbol{C}^{-1}(\mathsf{E}\left[\boldsymbol{b}\boldsymbol{b}^T\right] - \mathsf{E}\left[\boldsymbol{b}\right]\mathsf{E}\left[\boldsymbol{b}^T\right])\boldsymbol{C}^{-T}$ respectively. Substituting Equation (5.6) into the expression for $\boldsymbol{C}$, we obtain

$$\begin{aligned}
\boldsymbol{C} &= \sum_{n=M}^{N-M-1}\sum_{l=-M}^{M} s(n+l)s^*(n-l)\left[\frac{\partial^2}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}^T}e^{-j2\tilde{\omega}(n;\boldsymbol{\vartheta})l}\right]_{\boldsymbol{\vartheta}_0} \\
&= \sum_{n=M}^{N-M-1}\sum_{l=-M}^{M} s(n+l)s^*(n-l)e^{-j2\tilde{\omega}(n;\boldsymbol{\vartheta}_0)l} \\
&\quad\times\left[-j2l\frac{\partial^2\tilde{\omega}(n;\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}^T} + 4l^2\frac{\partial\tilde{\omega}(n;\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}}\frac{\partial\tilde{\omega}(n;\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}^T}\right]_{\boldsymbol{\vartheta}_0} \\
&= |A|^2\sum_{n=M}^{N-M-1}\sum_{l=-M}^{M}\left[-j2l\frac{\partial^2\tilde{\omega}(n;\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}^T} + 4l^2\frac{\partial\tilde{\omega}(n;\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}}\frac{\partial\tilde{\omega}(n;\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}^T}\right]_{\boldsymbol{\vartheta}_0}
\end{aligned}$$

where the final simplification above is made using Assumption 5. Expanding the expression for $\boldsymbol{b}$ using Equation (5.6) we obtain

$$\begin{aligned}\boldsymbol{b} &= -j2\sum_{n=M}^{N-M-1}\sum_{l=-M}^{M} e^{-j2\tilde{\omega}(n;\boldsymbol{\vartheta}_0)l}\, l\left[\frac{\partial\tilde{\omega}(n;\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}}\right]_{\boldsymbol{\vartheta}_0} \\ &\quad\times\left[s(n+l)V^*(n-l)+s^*(n-l)V(n+l)+V(n+l)V^*(n-l)\right].\end{aligned}$$

Using the assumption of AWGN,

$$\mathsf{E}\left[\boldsymbol{b}\right] = -j2\sigma_V^2\sum_{n=M}^{N-M-1}\sum_{l=-M}^{M}\delta(l)le^{-j2\tilde{\omega}(n;\boldsymbol{\vartheta}_0)l}\left[\frac{\partial\tilde{\omega}(n;\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}}\right]_{\boldsymbol{\vartheta}_0} = 0$$

which means that, to a first order approximation, $\mu_{\delta\boldsymbol{\vartheta}} = 0$ and the estimator is unbiased. The covariance matrix of $\boldsymbol{b}$ is given by

$$\begin{aligned}\mathsf{E}\left[\boldsymbol{b}\boldsymbol{b}^H\right] &= 4\sum_{n=M}^{N-M-1}\sum_{m=M}^{N-M-1}\sum_{l=-M}^{M}\sum_{k=-M}^{M} lke^{-j2[\tilde{\omega}(n;\boldsymbol{\vartheta}_0)l-\tilde{\omega}(m;\boldsymbol{\vartheta}_0)k]} \\ &\quad\times\left[\frac{\partial\tilde{\omega}(n;\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}}\frac{\partial\tilde{\omega}(m;\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}^T}\right]_{\boldsymbol{\vartheta}_0} \\ &\quad\times\Big[s(n+l)s^*(m+k)\mathsf{E}\left[V^*(n-l)V(m-k)\right] \\ &\quad+s^*(n-l)s(m-k)\mathsf{E}\left[V(n+l)V^*(m+k)\right] \\ &\quad+\mathsf{E}\left[V(n+l)V^*(n-l)V(m-k)V^*(m+k)\right]\Big],\end{aligned}$$

which, using Assumption 5 and the noise properties, may be simplified as

$$\begin{aligned}\mathsf{E}\left[\boldsymbol{b}\boldsymbol{b}^H\right] &= 4\sum_{n=M}^{N-M-1}\sum_{m=M}^{N-M-1}\sum_{l=-M}^{M}\sum_{k=-M}^{M}\left[\frac{\partial\tilde{\omega}(n;\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}}\frac{\partial\tilde{\omega}(m;\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}^T}\right]_{\boldsymbol{\vartheta}_0} lk \\ &\quad\times\Big[|A|^2\sigma_V^2[\delta(n-m-l+k)+\delta(n-m+l-k)] \\ &\quad+\sigma_V^4[\delta(l)\delta(k)+\delta(n-m+l-k)\delta(n-m-l+k)]\Big].\end{aligned}$$

Exploiting symmetry in the summations, and the fact that $\sum_{l=-M}^{M} l^2 = \frac{1}{3}M(M+1)(2M+1)$, we obtain the final simplification given in Equation (5.17), where $\boldsymbol{B} \triangleq \mathsf{E}\left[\boldsymbol{b}\boldsymbol{b}^H\right]$ ■

5.C The Cramér-Rao Bound for FM signals

In this section we derive a general form of the Cramér-Rao Bound (CRB) for estimating the parameter vector $\boldsymbol{\vartheta} \in \mathbb{R}^{p\times 1}$ from observations $\{X(n) = s(n) + V(n)\}_{n=0}^{N-1}$ where $s(n) = Ae^{j\psi(n;\boldsymbol{\vartheta})}$ and $V(n)$ is a complex white Gaussian noise sequence of variance σ_V^2. The full vector of unknown parameters is $\boldsymbol{\eta} = [\boldsymbol{\vartheta}^T, |A|, \angle A,$

$\sigma_V^2]^T$. We organize the observations and signal samples into vectors $\boldsymbol{X} = [X(0), \ldots, X(N-1)]^T$ and $\boldsymbol{s} = (s(0), \ldots, s(N-1))^T$ respectively. The CRB is given by the inverse of the Fisher information matrix (FIM); $\boldsymbol{J}$. Given a Gaussian distributed random vector $\boldsymbol{x}$, the FIM can be calculated according to [107]:

$$J_{pq} = \mathsf{Tr}\left[\boldsymbol{\Gamma}^{-1}\frac{\partial\boldsymbol{\Gamma}}{\partial\eta_p}\boldsymbol{\Gamma}^{-1}\frac{\partial\boldsymbol{\Gamma}}{\partial\eta_q}\right] + 2\mathsf{Re}\left[\frac{\partial\boldsymbol{\mu}^H}{\partial\eta_p}\boldsymbol{\Gamma}^{-1}\frac{\partial\boldsymbol{\mu}^H}{\partial\eta_p}\right], \tag{5.32}$$

where $\boldsymbol{\Gamma}$ and$\boldsymbol{\mu}$ correspond to the covariance matrix and mean vector of $\boldsymbol{X}$ respectively. Clearly $\boldsymbol{\Gamma} = \sigma_V^2\boldsymbol{I}$ and $\boldsymbol{\mu} = \boldsymbol{s}$ in this case. We first find the partial derivatives of $\boldsymbol{s}$ with respect to $\boldsymbol{\eta}$ as

$$\frac{\partial\boldsymbol{s}}{\partial\boldsymbol{\eta}^T} = \left[j\frac{\partial\boldsymbol{\psi}(\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}_p} \odot \boldsymbol{s}\ ,\ |A|^{-1}\boldsymbol{s}\ ,\ j\boldsymbol{s}\ ,\ \mathbf{0}_N\right] \tag{5.33}$$

where $\mathbf{0}_N$ denotes a length N column vector of zeros. In Equation (5.33) we have used the definition $\boldsymbol{\psi}(\boldsymbol{\vartheta}) = (\psi(0;\boldsymbol{\vartheta}), \ldots, \psi(N-1;\boldsymbol{\vartheta}))^T$. We wish to construct the FIM matrix in block form

$$\boldsymbol{J} = \begin{bmatrix} \boldsymbol{J}_{11} & \boldsymbol{J}_{12} \\ \boldsymbol{J}_{21} & \boldsymbol{J}_{22} \end{bmatrix}$$

such that the entries concerned with $\boldsymbol{\vartheta}$ are contained in $\boldsymbol{J}_{11}$. Using (5.32) and the partial derivatives of $\boldsymbol{s}$ from Equation (5.33) , we find

$$\begin{aligned} \boldsymbol{J}_{11} &= 2|A|^2/\sigma_V^2\boldsymbol{Q}^T\boldsymbol{Q} \\ \boldsymbol{J}_{12} &= \boldsymbol{J}_{21}^T = [\ \mathbf{0}_p\ ,\ 2|A|^2/\sigma_V^2\boldsymbol{Q}^T\mathbf{1}_N\ ,\ \mathbf{0}_p] \\ \boldsymbol{J}_{22} &= \mathsf{Diag}[\ 2N/\sigma_V^2\ ,\ 2N|A|^2/\sigma_V^2\ ,\ N/\sigma_V^4\ ] \end{aligned}$$

where $\boldsymbol{Q} = \partial\boldsymbol{\psi}(\boldsymbol{\vartheta})/\partial\boldsymbol{\vartheta}^T$ and $\mathbf{1}_N$ is a length N column vector of ones. By use of the partitioned matrix inversion identity we may find the CRB with respect to $\boldsymbol{\vartheta}$ as

$$\begin{aligned} \boldsymbol{C}(\boldsymbol{\vartheta}) &= (\boldsymbol{J}_{11} - \boldsymbol{J}_{12}\boldsymbol{J}_{22}^{-1}\boldsymbol{J}_{21})^{-1} \\ &= \frac{N}{2SNR_{in}}(\boldsymbol{Q}^T(N\boldsymbol{I} - \mathbf{1}_N^T\mathbf{1}_N)\boldsymbol{Q})^{-1} \end{aligned}$$

where $SNR_{in} = |A|^2/\sigma_V^2$. ■

Chapter 6

Micro-Doppler Signature Estimation

In this chapter, the estimation of micro-Doppler TF signatures is considered. The problem and model are first reviewed in Section 6.1. Optimal estimation of micro-Doppler parameters is then considered in Section 6.2. The PWHT estimator introduced in the previous chapter is discussed in Section 6.3, where the problem of estimation bias is considered in detail. Finally, experimental data is used to validate the proposed estimation strategy.

6.1 The micro-Doppler phenomenon

The Doppler effect was discussed in Chapter 1 as a motivating example for the analysis of nonstationary signals. A well known example occurs in radar systems, where a transmitted electromagnetic signal is reflected by a target back toward the radar system [101]. If the target is at a distance r_0 and the radar signal has wavelength λ, then the total phase change in the two-way propagation is given by

$$\psi = 2\pi \times \frac{2r_0}{\lambda} = 4\pi r_0/\lambda.$$

If the target is in motion relative to the radar, the distance is time-varying which in turn results in a time-varying phase. In the simple case of a target moving at a constant radial velocity, v, in the radar line of sight (LOS), the distance is changing according to $r(t) = r_0 + v_0 t$ and the phase is given by

$$\psi(t) = 4\pi \left(\frac{r_0}{\lambda} + \frac{v_0}{\lambda} t \right) = 4\pi \left(\frac{r_0}{\lambda} + \frac{v_0 f_c}{c} t \right), \tag{6.1}$$

where f_c is the radar carrier frequency and c is the speed of electromagnetic wave propagation in the medium. Differentiating Equation (6.1) with respect to time gives the instantaneous frequency

$$\tilde{\omega}(t) = \frac{d\psi(t)}{dt} = 2\pi \times \frac{2v_0}{c} f_c = 2\pi f_d, \tag{6.2}$$

where f_d corresponds to the Doppler frequency shift.

For targets with more complex motion, or time-varying velocity, the resulting frequency shift may also be time-varying. In certain cases, such as when the target endures simple harmonic motion such as vibration or rotation, the phase modulation can be sinusoidal [16,75,117]. These kinds of motions arise in a number of applications such as microphone calibration [110], medical imaging [120], radar target identification [16] and indoor radar imaging [4,100]. For instance, if the radar signal is back-scattered from the moving parts of the compressor and blade assembly of a jet airplane, this leads to the Jet Engine Modulation (JEM) effect [16] which can well modeled as sinusoidal FM. Another example is a radar system observing a walking person, where the swinging action of the person's arms and hands induces a typical micro-Doppler effect in the back-scattered waveform [39,41,42].

In the following, the signal model for the micro-Doppler signal is derived. In later sections, the problem of estimating micro-Doppler parameters shall be addressed.

6.1.1 Point scatterer model

In this section, the micro-Doppler signal model and the corresponding spectrum are derived. We consider a radar system located at the origin of a coordinate system which is observing a target (point scatterer) at polar coordinate position (r_0, θ, ϕ), as depicted in Figure 6.1.

It is assumed that the point scatterer is undergoing simple harmonic motion and the coordinate system is aligned such that the motion is parallel to the x-axis. The position of the scatterer at time t is therefore given by:

$$x(t) = d_x - d \sin(\omega_m t + \phi),$$

where $d_x = r_0 \sin\theta \sin\phi$. Due to the target motion, the range from the target to the radar system is time-varying according to:

$$r(t) = \sqrt{x^2(t) + (r_0 \cos\theta \sin\phi)^2 + (r_0 \sin\phi)^2}.$$

Assuming the vibration amplitude of the target is small compared to the distance to the radar, i.e. $d << r_0$, the range may be approximated as [102]:

$$r(t) \approx r_0 - (d \sin\theta \sin\phi) \sin(\omega_m t + \phi). \tag{6.3}$$

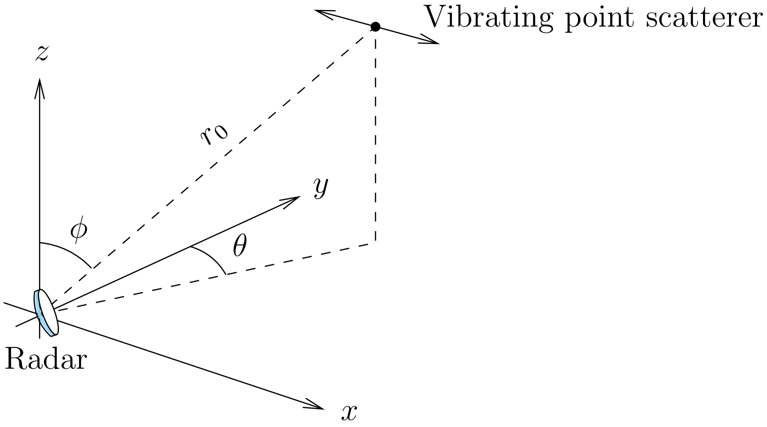


Figure 6.1:

The above equation has the following interpretation: the range is varying according to the projection of the target motion on to the LOS of the radar. Clearly, if $\theta = 0$, i.e. the target lies in the $y - z$ plane, then the effect of the oscillation is perpendicular to the LOS and will not be observed by the radar. Conversely, a target motion parallel to the LOS will have the most pronounced effect (cause the most significant phase change) in the radar return. The range 'deviation' of the target is now defined as:

$$\delta r(t) \triangleq r(t) - r_0 = -\tilde{d} \sin(\omega_m t + \phi),$$

where $\tilde{d} = (d \sin\theta \sin\phi)$. The target range deviation results in a modulation of the phase, $\psi(t)$, in the radar return signal, according to:

$$\psi(t) = 2\frac{2\pi}{\lambda}\delta r(t) = B \sin(\omega_m t + \phi), \tag{6.4}$$

where $B \triangleq -4\pi\tilde{d}/\lambda$ is the modulation index and ω_m is the angular modulation frequency. We note that, although Equation (6.4) was obtained under the assumption of vibrational motion, the same result follows from rotational motion[1].

In addition to a "local" harmonic motion, a target may simultaneously have "global" motion with respect to the radar system. This global motion may be a linear velocity or acceleration, which also results in a phase modulation (Doppler). We consider the return from a single point-scatterer, exhibiting both Doppler and micro-Doppler phenomena, to be modeled as follows:

$$s(t) = Ae^{j\psi_d(t)}e^{j\psi_m(t)} \tag{6.5}$$

where A represents a complex random amplitude, and $\psi_d(t)$ and $\psi_m(t)$ represent the phase modulation of the signal due to Doppler and micro-Doppler effects,

---

[1] See e.g. [40] for a more general model of the target motion.

respectively. We assume that the Doppler phase term can be expressed as a (generally low-order) polynomial in time of order q [97]:

$$\psi_d(t) = \sum_{k=1}^{q} a_k t^k. \tag{6.6}$$

For example, if $q = 1$ then Equation (6.6) models the (constant) Doppler frequency shift induced due to a target moving at constant radial velocity with respect to the receiver. If $q = 2$ then a_2 is determined by the radial acceleration of the target, etc.

6.1.2 The micro-Doppler spectrum

The spectrum of the micro-Doppler signals is well known in the literature (see e.g. [68, 75]). Here we give a derivation of the Fourier transform of the micro-Doppler signal based on Equation (6.4).

Fourier transform We define the micro-Doppler component of the signal as $s_m(t) = e^{j\psi_m(t)}$. The Fourier transform of $s_m(t)$ is given by

$$S_m(j\Omega) = \sum_{k=-\infty}^{\infty} J_k(B) e^{j\phi k} \delta(\Omega - w_m k), \tag{6.7}$$

where $\delta(\cdot)$ denotes Dirac's delta function and $J_n(\cdot)$ denotes the nth Bessel function of the first kind. From Equation (6.7) we see that $S_m(j\Omega)$ is made up of harmonic components, spaced ω_m rad/sec apart, each weighted by a Bessel function dependent on B. Therefore two parameters of interest, namely ω_m and B, may potentially be inferred from the spacing and amplitudes of peaks in $S_m(j\Omega)$. A number of approaches for estimation of the micro-Doppler parameters have been proposed which exploit these properties (see e.g. [75] and the references therein).

Proof In the proof of Equation (6.7) we have to make use of the identity

$$e^{jx \sin\theta} = \sum_{n=-\infty}^{\infty} J_n(x) e^{j\theta}$$

The Fourier transform of $s(t)$ is derived as follows:

$$\begin{aligned}
S_m(j\Omega) &= \int_{\mathbb{R}} s_m(t)e^{-j\Omega t}dt \\
&= \int_{\mathbb{R}} e^{jB\sin(w_m t+\phi)}e^{-j\Omega t}dt \\
&= \int_{\mathbb{R}} \left[\sum_{k=-\infty}^{\infty} J_k(B)e^{j(w_m t+\phi)k}\right] e^{-j\Omega t}dt \\
&= \sum_{k=-\infty}^{\infty} J_k(B)e^{j\phi k}\int_{\mathbb{R}} e^{-j(\Omega-w_m k)t}dt \\
S_m(j\Omega) &= \sum_{k=-\infty}^{\infty} J_k(B)e^{j\phi k}\delta(\Omega-w_m k).
\end{aligned}$$

6.2 Parameter Estimation in Gaussian noise

For a signal modelled by Equation (6.5) observed in additive Gaussian noise, we derive the Maximum Likelihood (ML) estimator and the Cramér-Rao lower bound (CRB) on the estimator variance for the parameters of interest. We assume there are N observations of the signal in additive noise, modeled according to:

$$X(n) = s(n) + V(n); \quad n = 1, \ldots, N, \tag{6.8}$$

with $s(n)$ being a sampled version of the signal defined in Equation (6.5) and $V(n)$ a complex circular white Gaussian noise process of variance σ_V^2. The vector of unknown parameters is

$$\boldsymbol{\eta} = [B, \omega_m, \phi, a_1, \ldots, a_q, A_0, \theta_0, \sigma_V^2]^{\mathrm{T}} \tag{6.9}$$

where $A_0 \in \mathbb{R}^+$ and $\theta_0 \in [0, 2\pi)$ are the magnitude and phase of A from Equation (6.5) respectively. We organize the observations into a vector $\boldsymbol{X} = (X(0), \ldots, X(N-1))'$ and define $\boldsymbol{s}$ and $\boldsymbol{V}$ comfortably with $\boldsymbol{X}$. The distribution of the data is therefore compactly described as $\boldsymbol{X} \sim \mathcal{N}^C(\boldsymbol{\mu}, \boldsymbol{\Gamma})$ with mean vector $\boldsymbol{\mu} = \boldsymbol{s}$ and the covariance matrix $\boldsymbol{\Gamma} = \sigma_V^2\boldsymbol{I}$.

6.2.1 The ML estimator

The log-likelihood function of the observations $\boldsymbol{x} = [\boldsymbol{x}(0), \ldots, \boldsymbol{x}(N-1)]^{\mathrm{T}}$ with respect to the parameters is given by

$$L(\boldsymbol{\eta}, \boldsymbol{x}) = -\frac{N}{2}(\ln 2\pi + 2\ln\sigma_V) - \frac{1}{2\sigma_V^2}(\boldsymbol{x}-\boldsymbol{s})^H(\boldsymbol{x}-\boldsymbol{s}). \tag{6.10}$$

ML estimation of $\boldsymbol{\eta}$ is achieved by maximization of $L(\eta, \boldsymbol{x})$ with respect to $\boldsymbol{\eta}$ for a given vector $\boldsymbol{x}$ of observations. If one is only interested in the signal parameters $\boldsymbol{\eta}_s = (B, \omega_m, \phi, a_1, \ldots, a_q, A_0, \theta_0)'$ then the ML estimate of $\boldsymbol{\eta}_s$ is obtained via minimization of a cost function:

$$\hat{\boldsymbol{\eta}}_s = \arg\min_{\boldsymbol{\eta}_s} (\boldsymbol{x} - \boldsymbol{s}(\boldsymbol{\eta}_s))^H (\boldsymbol{x} - \boldsymbol{s}(\boldsymbol{\eta}_s)). \tag{6.11}$$

Since this cost function is non-linear and non-convex in $\boldsymbol{\eta}_s$, the optimization step required in Equation (6.11) is computationally expensive, and often relies on good initialization of the optimization algorithm in order to reach the global minimum. A number of sub-optimal estimators for the micro-Doppler and Doppler parameters exist [65,75,117], which might be used for obtaining initial estimates. An alternative approach based on the model in Equation (6.5) is proposed and discussed in Appendix 6.B.

The ML estimator discussed above leads to asymptotically efficient estimation, in the case of a single micro-Doppler signal in additive white Gaussian noise. However, applying the optimal estimator in the multicomponent case, where a number of micro-Doppler signatures may be simultaneously present in the recorded data, quickly becomes intractable, as the computational complexity rises exponentially with the number of unknown parameters. If there is any amplitude modulation present in the signal, this must also be appropriately modeled for derivation of the optimal estimator. An alternative approach, suitable for estimating the micro-Doppler parameters of multicomponent signals, which is insensitive to amplitude modulations, is the PWHT presented in Chapter 5. The application of the PWHT to micro-Doppler signature estimation shall be investigated in Section 6.3.

6.2.2 The CRB

Let $\hat{\boldsymbol{\eta}}$ be an unbiased estimator of the parameter vector $\boldsymbol{\eta}$. The mean squared error (MSE) when estimating parameter η_i is lower bounded according to $\mathsf{E}\left[(\hat{\eta}_i - \eta_i)^2\right] \leq [\boldsymbol{J}^{-1}]_{ii}$, for $i = 1, \ldots, p$, where

$$\boldsymbol{J} = \begin{bmatrix} \boldsymbol{J}_{mm} & \boldsymbol{J}_{md} & \boldsymbol{J}_{mA} & \mathbf{0} \\ \boldsymbol{J}_{dm} & \boldsymbol{J}_{dd} & \boldsymbol{J}_{dA} & \mathbf{0} \\ \boldsymbol{J}_{Am} & \boldsymbol{J}_{Ad} & \boldsymbol{J}_{AA} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & N/\sigma_V^4 \end{bmatrix} \tag{6.12}$$

with

$$\begin{aligned}
\boldsymbol{J}_{mm} &= K\begin{bmatrix} \boldsymbol{\varphi}'_m\boldsymbol{\varphi}_m & B\boldsymbol{\varphi}'_m(\boldsymbol{\psi}_m \odot \boldsymbol{t}^{(1)}) & B\boldsymbol{\varphi}'_m\boldsymbol{\psi}_m \\ B\boldsymbol{\varphi}'_m(\boldsymbol{\psi}_m \odot \boldsymbol{t}^{(1)}) & B^2(\boldsymbol{\psi} \odot \boldsymbol{\psi}_m)'\boldsymbol{t}^{(2)} & B^2(\boldsymbol{\psi}_m \odot \boldsymbol{\psi}_m)'\boldsymbol{t}^{(1)} \\ B\boldsymbol{\varphi}'_m\boldsymbol{\psi}_m & B^2(\boldsymbol{\psi}_m \odot \boldsymbol{\psi}_m)'\boldsymbol{t}^{(1)} & B^2\boldsymbol{\psi}'_m\boldsymbol{\psi}_m \end{bmatrix} \\
\boldsymbol{J}_{dd} &= K\boldsymbol{T}'\boldsymbol{T} \\
\boldsymbol{J}_{AA} &= K\begin{bmatrix} A_0^{-2} & 0 \\ 0 & 1 \end{bmatrix} \\
\boldsymbol{J}_{md} &= \boldsymbol{J}'_{dm} = K\begin{bmatrix} \boldsymbol{\varphi}'_m\boldsymbol{T} \\ B(\boldsymbol{\psi}_m \odot \boldsymbol{t}^{(1)})'\boldsymbol{T} \\ B\boldsymbol{\psi}'_m\boldsymbol{T} \end{bmatrix} \\
\boldsymbol{J}_{mA} &= \boldsymbol{J}'_{Am} = K\begin{bmatrix} 0 & \boldsymbol{\varphi}'_m\boldsymbol{1} \\ 0 & B(\boldsymbol{\psi}_m \odot \boldsymbol{t}^{(1)})'\boldsymbol{1} \\ 0 & B\boldsymbol{\psi}'_m\boldsymbol{1} \end{bmatrix} \\
\boldsymbol{J}_{dA} &= \boldsymbol{J}'_{Ad} = K\begin{bmatrix} 0 & \boldsymbol{T}'\boldsymbol{1} \end{bmatrix}
\end{aligned}$$

where $K = 2NA_0^2/\sigma_V^2 = 2N\mathrm{SNR}$ and the following definitions have been used:

$$\begin{aligned}
\boldsymbol{\varphi}_m &= (\sin(\omega_m t_1 + \phi), \ldots, \sin(\omega_m t_N + \phi))' \\
\boldsymbol{\psi}_m &= (\cos(\omega_m t_1 + \phi), \ldots, \cos(\omega_m t_N + \phi))' \\
\boldsymbol{t}^{(i)} &= (t_1^i, \ldots, t_N^i)' \\
\boldsymbol{T} &= (\boldsymbol{t}^{(1)}, \ldots, \boldsymbol{t}^{(q)})'.
\end{aligned} \tag{6.13}$$

The derivation of the CRB is given in Appendix 6.A.

6.3 Parameter estimation using the PWHT

A suboptimal approach, suitable for estimation of multicomponent micro-Doppler signatures, has been introduced in the previous chapter. It was demonstrated that the PWHT allows one to estimate the parameters of the frequency modulation with accuracy close to the CRB, in a computationally efficient manner. The use of the PWHT, however, can result in a bias for highly nonlinear FM signals. For the case of micro-Doppler signals which exhibit a sinusoidal FM, this bias may be significant, especially for estimation of the FM amplitude parameter B. In the following section the issue of bias for the PWHT estimator is discussed in more detail. Application of the PWHT to estimation of micro-Doppler signatures is then demonstrated using experimental data in Section 6.3.2.

6.3.1 Bias of the PWHT estimator

The discrete-time PWHT of the random sample $\{X(n)\}_{n=0}^{N-1}$ is defined according to:

$$P_X(\boldsymbol{\vartheta}) = \sum_{n=M}^{(N-M-1)} \hat{D}_{XX}(n, \omega(n, \boldsymbol{\vartheta})), \tag{6.14}$$

where $\hat{D}_{XX}(n, \omega)$ in the above denotes the discrete-time PWVD estimator:

$$\hat{D}_{XX}(n, \omega) = \sum_{l=-M}^{M} X(n+l)X^*(n-l)e^{j2\omega l}.$$

In the following we shall assume that the Doppler phase component $\psi_d(t)$ of Equation (6.5) is zero or has been estimated and removed, so that the IF function is given by:

$$\tilde{\omega}(t; \boldsymbol{\vartheta}) = B \sin(\omega_m t + \phi) \tag{6.15}$$

and the FM parameters are $\boldsymbol{\vartheta} = [B, \omega_m, \phi]^{\mathrm{T}}$. We note that the definition given above is consistent with the derivation in Section 6.1.1 for $B = \frac{4\pi}{\lambda}\tilde{d}\omega_m$.

The estimator of the IF parameters for a micro-Doppler signal is obtained as the location of the maximum of $P_X(\boldsymbol{\vartheta})$ with respect to $\boldsymbol{\vartheta}$. The maximum occurs when the parameter $\boldsymbol{\vartheta}$ describes a trajectory in the TF plane which passes along the highest "ridge" of the PWVD. Assuming the signal IF is approximately linear within each time window of length $L = 2M+1$ samples, this ridge coincides approximately with the signal IF. However, if the signal IF is highly nonlinear within the window duration, the ridge of the PWVD may deviate significantly from the true IF, resulting in biased estimation of the FM parameters using Equation (6.14).

To illustrate how the window length, L, of the PWVD can effect the bias of the PWHT, an example is given. In Figure 6.2, the true signal IF of a micro-Doppler signal is plotted along with the peak of the PWVD, computed using various window lengths. As can be seen, an increased window length results in a ridge which deviates from the true IF at the peaks and troughs of the sinusoidal FM (at which points the IF is the most nonlinear). As the peak of the PWHT corresponds to the trajectory along the highest ridge of the PWVD, the estimation of the amplitude, B, of the sinusoidal FM will be certainly biased. It is noted that the estimation of the frequency ω_m and phase ϕ will not be biased as these parameters are determined by the zero-crossings of the FM trajectory, at which points the IF is close to linear.

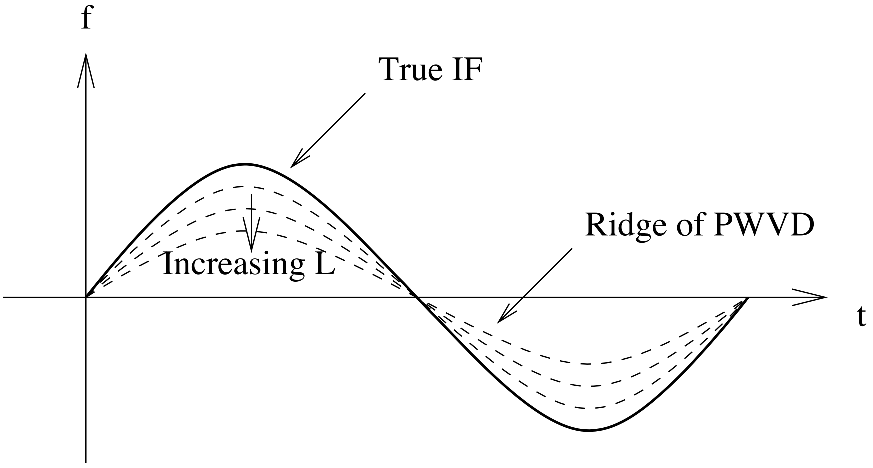


Figure 6.2: The true signal IF of a micro-Doppler signal (−) and the highest ridge of the PWVD (−−) for various values of L.

As discussed in the preceding chapter, the location of the peak of Equation (6.14) occurs at the point $\boldsymbol{\vartheta}_s + \delta\boldsymbol{\vartheta}$, where $\delta\boldsymbol{\vartheta}$ is a zero-mean random variable and $\boldsymbol{\vartheta}_s$ is the location of the maximum of $P_s(\boldsymbol{\vartheta})$. Let us assume that the signal $s(t)$ has the micro-Doppler IF model given in Equation (6.15), with parameter values $\boldsymbol{\vartheta}_0 = [B_0, \omega_{m,0}, \phi_0]^{\mathrm{T}}$. The peak of $P_s(\boldsymbol{\vartheta})$ will therefore occur at $\boldsymbol{\vartheta}_s = \boldsymbol{\vartheta}_0 + [\delta B, 0, 0]^{\mathrm{T}}$, where δB corresponds to the difference between B_0 and the peak of $D_{ss}(n_0, \omega)$ where $n_0 = (\pi/2 - \phi_0)/\omega_m$ (illustrated in Figure 6.3). Determining the location of the maximum of $D_{ss}(n_0, \omega)$ will therefore allow one to derive and expression for the bias δB.

An approximate expression for δB may be derived, for example, using a Taylor series expansion, as suggested in [118]. Following this approach, a first order approximation requires one to solve the equation

$$\delta B \approx - \left[\left(\frac{\delta D_{ss}(n_0, \omega)}{\delta \omega} \right) \left(\frac{\delta^2 D_{ss}(n_0, \omega)}{\delta \omega^2} \right)^{-1} \right]_{\omega = B_0} \tag{6.16}$$

Substituting the model $s(n) = \exp[\mathrm{j}\psi_s(n)]$ into Equation (6.16), where $\psi_s(n) = -B_0/\omega_{m,0} \cos(\omega_{m,0} + \phi_0)$ consistent with Equation (6.15), yields

$$D_s(n_0, \omega) = \sum_{l=-M}^{M} \mathrm{e}^{-2\mathrm{j}[\omega l - B_0/\omega_{m,0} \sin(\omega_{m,0} l)]}. \tag{6.17}$$

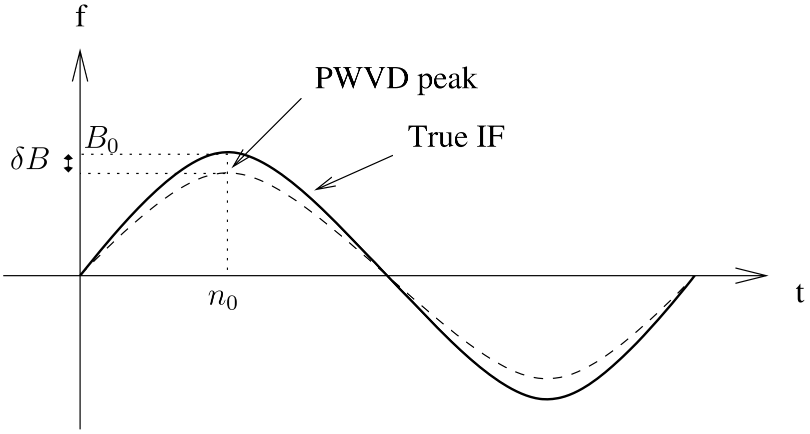


Figure 6.3: For a micro-Doppler amplitude B_0, the location of the peak of the PWVD occurs at a frequency $B_0 + \delta B$.

Substituting Equation (6.17) into Equation (6.16) yields the approximate bias expression[2]:

$$\delta B \approx \frac{\mathrm{j}\sum_{l=-M}^{M} l \exp\{-\mathrm{j}2B_0[l - \sin(\omega_{m,0}l)/\omega_{m,0}]\}}{2\sum_{l=-M}^{M} l^2 \exp\{-\mathrm{j}2B_0[l - \sin(\omega_{m,0}l)/\omega_{m,0}]\}}. \tag{6.18}$$

It is noted that the above expression is derived under the assumption that the PWVD peak occurs close to the true IF.

Given an estimate of $B_0 + \delta B$ from the PWHT of the observations, one would like to use the expression in Equation (6.18) to obtain an unbiased estimate of B_0. This is problematic for two reasons: Firstly, for large PWVD window lengths (where the PWHT variance is lower), the peak of the PWVD may not occur close enough to the true IF for the Taylor series approximation to be accurate. Secondly, the relationship between B_0 and δB in Equation (6.18) is highly nonlinear and does not allow one to easily solve for B_0.

A alternative approximation for δB is now proposed based on Equation (6.17). We note that the summation is bounded from above by $L = 2M + 1$. This bound may only be achieved if the angle of the exponential term is an integer multiple of 2π, which is only the case for a strictly linear FM signal. It is expected, however, that the maximum should be close to $2M + 1$ provided the PWVD window is small with respect to the micro-Doppler period. The peak frequency is therefore approximated by the minimizer of the angles $|\omega l - B_0/\omega_{m,0} \sin(\omega_{m,0} l)|$ for $l = -M, \dots, M$, with respect to ω. Using a least-squares formulation, we

---

[2] Despite the imaginary terms in Equation (6.18), the approximate bias expression remains real-valued due to the odd and even symmetry in the numerator and denominator expressions, respectively.

obtain

$$\begin{aligned}\delta B &\approx \left[\arg\min_{\omega} \sum_{l=-M}^{M} |\omega l - B_0/\omega_{m,0}\sin(\omega_{m,0} l)|^2\right] - B_0 \\ &= B_0\left[\frac{\sum_{l=-M}^{M} l\sin(\omega_{m,0} l)}{\omega_{m,0}\sum_{l=-M}^{M} l^2} - 1\right]. \end{aligned} \tag{6.19}$$

While the derivation of Equation (6.19) may be somewhat *ad hoc*, the justification is intuitively satisfying and empirical analysis has shown that it is more accurate the approximation based on the Taylor series expansion. Further, the relationship between δB and B_0 in Equation (6.19) is linear which allows one to easily perform bias correction. In Figure 6.4 the true value of δB for the PWVD is compared to the Taylor series (TS) approximation from Equation (6.18) and the proposed approximation from Equation (6.19). It can been seen that the proposed approximation is accurate even for L greater than half the micro-Doppler period, while the TS approximation breaks down for L greater than about 20% of the micro-Doppler period.

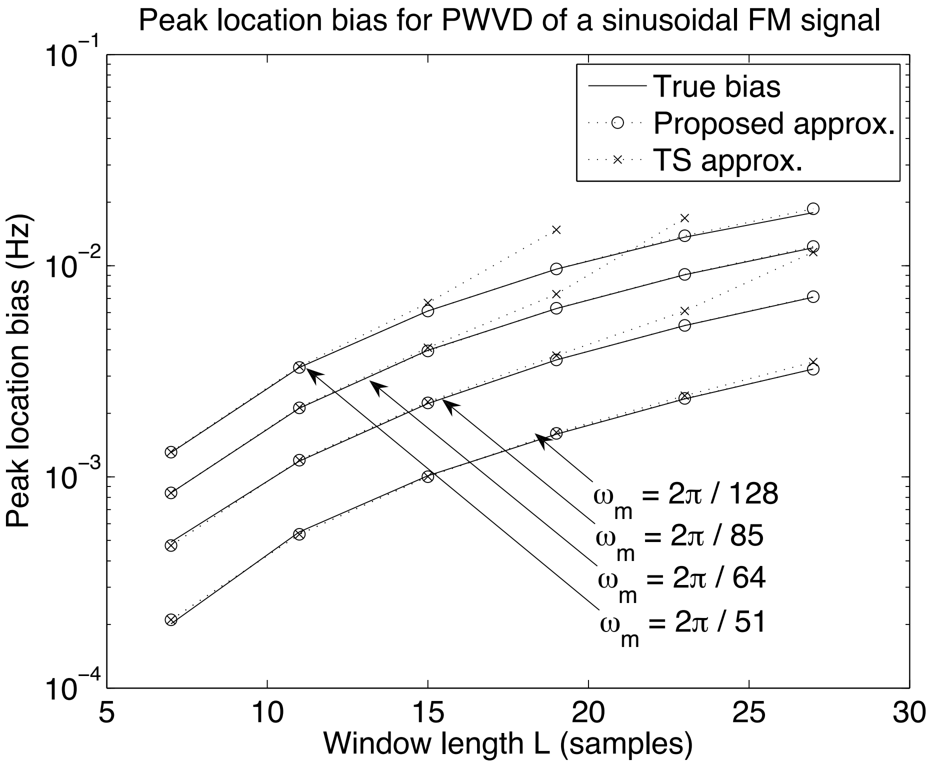


Figure 6.4: Comparison of the TS and proposed bias approximations for various values of the PWVD window length L and micro-Doppler frequency ω_m.

In order to improve the estimation accuracy using the PWHT, one may perform bias correction based on Equation (6.19), as outlined in Table 6.1. The idea is to first obtain estimates $\hat{\boldsymbol{\vartheta}} = [\hat{B}, \hat{\omega}_m, \hat{\phi}]^{\mathrm{T}}$ of the micro-Doppler FM parameters using the procedure given in Table 5.2. The estimates of ω_m and ϕ are then substituted into Equation (6.19) to form an estimate of the bias which is used to correct $\hat{B}$. Simulation results are shown in Figure 6.5 which indicate that the bias correction is successful. The simulation was conducted using micro-Doppler parameter values $\boldsymbol{\vartheta} = 2\pi[0.025,\ 2.5/N,\ 0.23]^{\mathrm{T}}$, with $N = 128$ observations and PWVD window lengths of $L = 7$ and $L = 17$, for SNR varying from 0 to 20 dB. 500 Monte Carlo simulations were averaged to obtain the estimator root mean-squared error (RMSE) shown in Figure 6.5, for estimation of B.

As indicated by the example in Figure 6.4, higher PWVD window lengths lead to higher bias. This is clear in the simulation results shown in Figure 6.5. Consider first the PWHT accuracy without bias correction: at low SNR, where variance is the dominant source of estimation error, the longer window length ($L = 17$) demonstrates the best accuracy. At high SNR, where the bias is the dominant source of estimation error, the PWHT estimator using the smaller window length ($L = 7$) has better accuracy. However, after applying bias correction according to Table 6.1, the accuracy of the PWHT becomes close to optimal at high SNR, even for $L = 17$.

1. Obtain an estimate, $\hat{\boldsymbol{\vartheta}} = [\hat{B}, \hat{\omega}_m, \hat{\phi}]^{\mathrm{T}}$, of the micro-Doppler parameters using the PWHT:

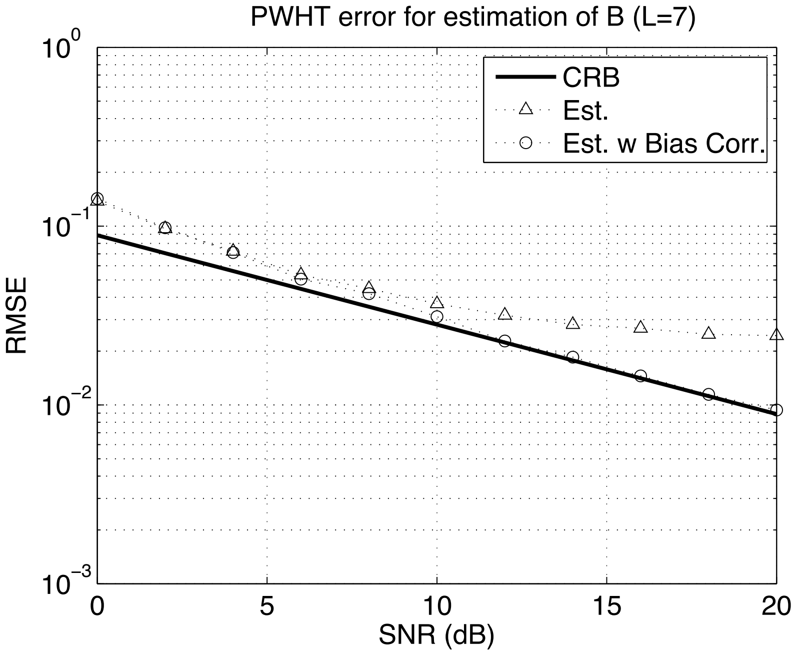
   $$\hat{\boldsymbol{\vartheta}} = \arg\min_{\boldsymbol{\vartheta}} P_x(\boldsymbol{\vartheta}),$$

   where the computationally efficient implementation proposed in Table 5.2 may be applied.

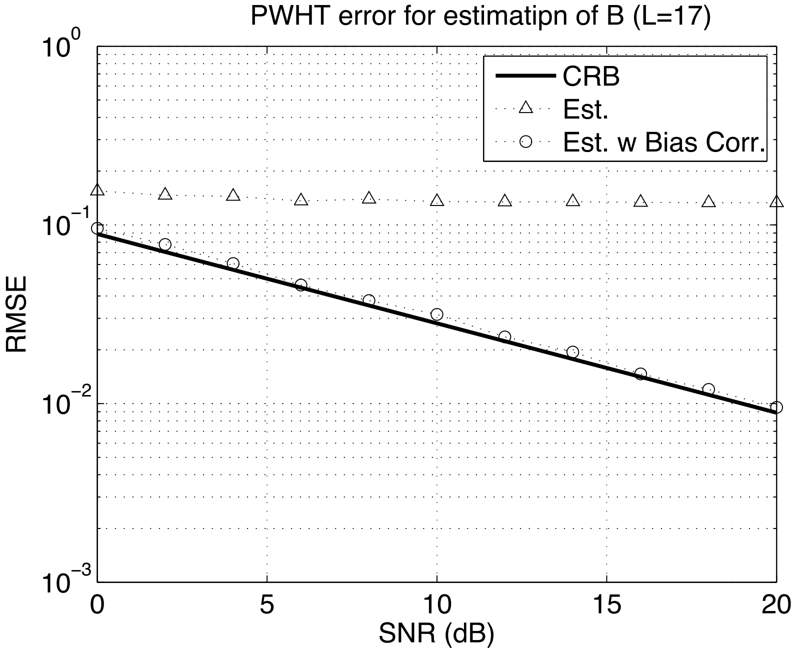
2. Correct the bias of the micro-Doppler amplitude according to:

   $$\hat{B} \leftarrow \hat{B}\left[\frac{\hat{\omega}_m \sum_{l=-M}^{M} l^2}{\sum_{l=-M}^{M} l \sin(\hat{\omega}_m l)}\right].$$

Table 6.1: Proposed bias correction for estimation of the micro-Doppler amplitude using the PWHT.



(a)



(b)

Figure 6.5: Estimation accuracy of the PWHT when using the proposed bias correction scheme. The PWVD window length used in $L = 7$ in (a) and $L = 17$ in (b).

6.3.2 Experimental results

The PWHT estimator is now demonstrated using experimental data which has been collected from a 24GHz radar system, observing a rotating fan. An image of the experimental setup is included as Figure 6.6. The rotational movement of the scatterer in this experiment results in a sinusoidal Doppler shift with respect to time (a micro-Doppler signature). To illustrate the estimation of multicomponent signatures, we apply the PWHT estimator to the data collected only from the in-phase baseband channel of the radar system. This effectively produces two TF "signatures" each π radians out of phase with the other.

The baseband radar signal was sampled at 1000 Hz and we have used an observation interval of 402 samples ($\sim$0.4 seconds) to estimate the micro-Doppler signatures. The proposed estimation strategy outlined in Table 5.2 has been followed, where the initial grid search was performed for $B \in 2\pi[0, 250]$ rad/s, $\phi \in [0, 2\pi)$ rad and $\omega_0 \in 2\pi[1, 10]$ rad/s, with 12, 10 and 6 samples along each parameter range respectively (720 total trajectories). In the initial search, $L = 31$ was used to calculate the PWHT, and in the final optimization step, $L = 71$.

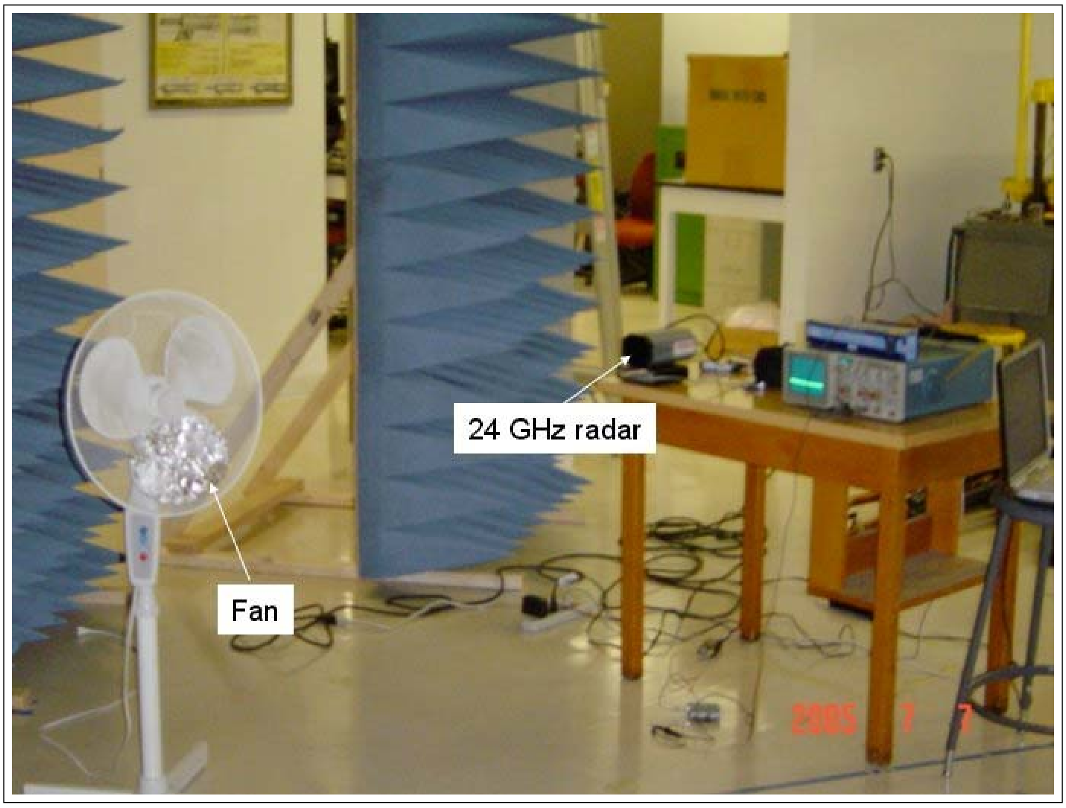


Figure 6.6: Experimental setup used for micro-Doppler radar data collection.

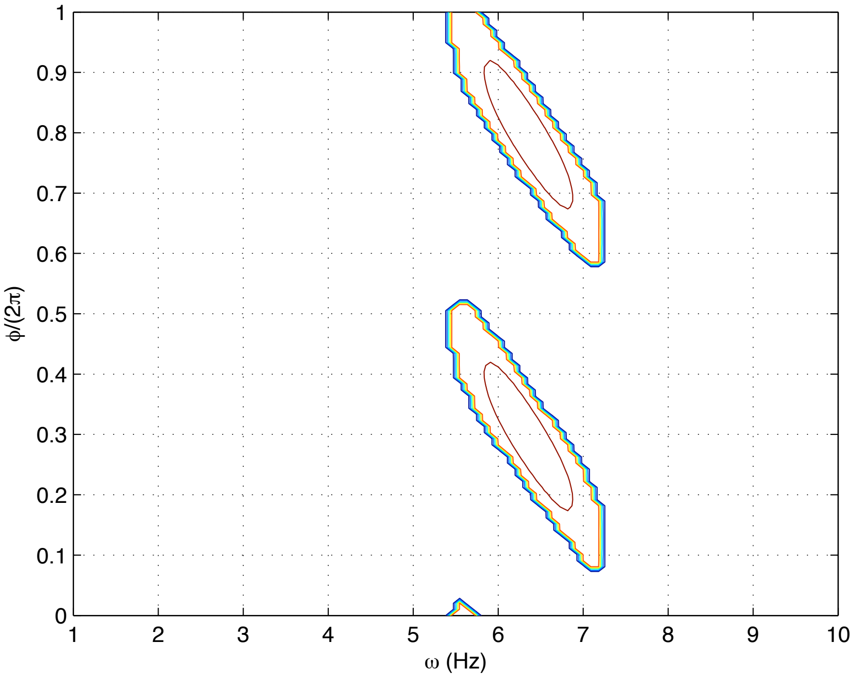
In Figure 6.7 we show the PWHT of the experimental data for the amplitudeslice of the parameter space at $B = 2\pi 16$ rad/s. The figure shows both cases of $L = 31$ and $L = 71$, which clearly illustrates the advantage of widening the main peak, achieved with the smaller window length, as discussed in Chapter 5. In Figure 6.8 we show the estimated micro-Doppler signatures overlaid on the PWVD of the data (for $L = 71$). In this figure we see both the initial grid search estimates and the final estimated signatures. It is observed that both the final estimated signatures overlap the TF signatures as expected, although the initial grid search yielded somewhat inaccurate results.

6.4 Conclusions

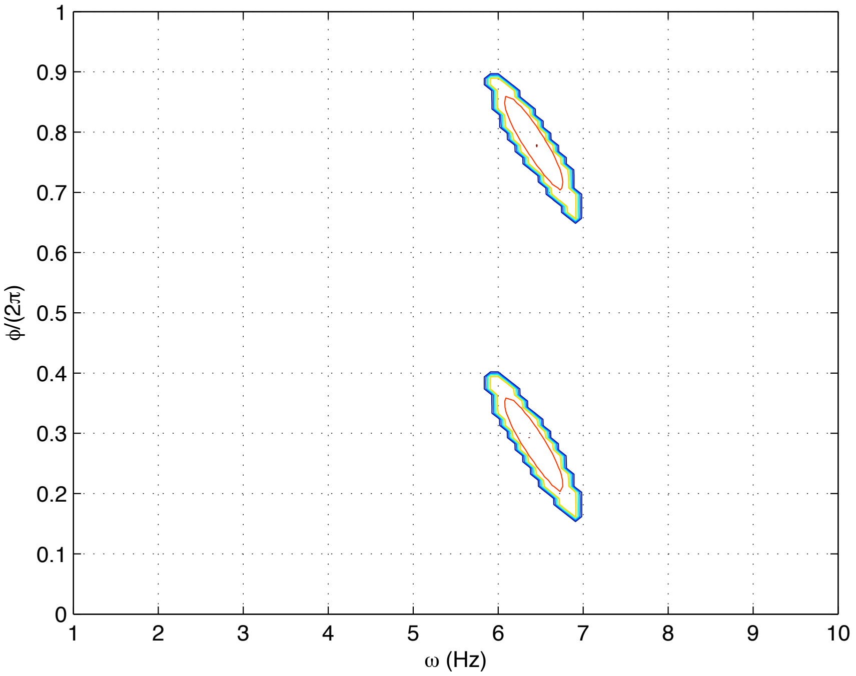
In this chapter the problem of micro-Doppler signature estimation has been considered. Due to rotational or vibrational movement of targets, the received signal of a radar system may exhibit a sinusoidal frequency modulation, known as micro-Doppler. The PWHT proposed in the previous chapter has been examined as a possible estimator for the sinusoidal FM parameters. It was shown that the estimation of the micro-Doppler amplitude parameter is biased using the PWHT, and a bias correction scheme was proposed. The PWHT was applied to experimental radar data and demonstrated to successfully estimate the TF signatures of the micro-Doppler signal resulting from a rotating fan.

6.5 Acknowledgments

I would like to sincerely thank Prof. Amin and his colleagues at the Wireless Communications and Positioning Lab in the Center for Advanced Communications at Villanova University, PA, for providing me with the experimental radar data.



(a)



(b)

Figure 6.7: PWHT for $L = 31$ (a) and $L = 71$ (b) of the experimental micro-Doppler data, evaluated for $B = 16$ Hz.

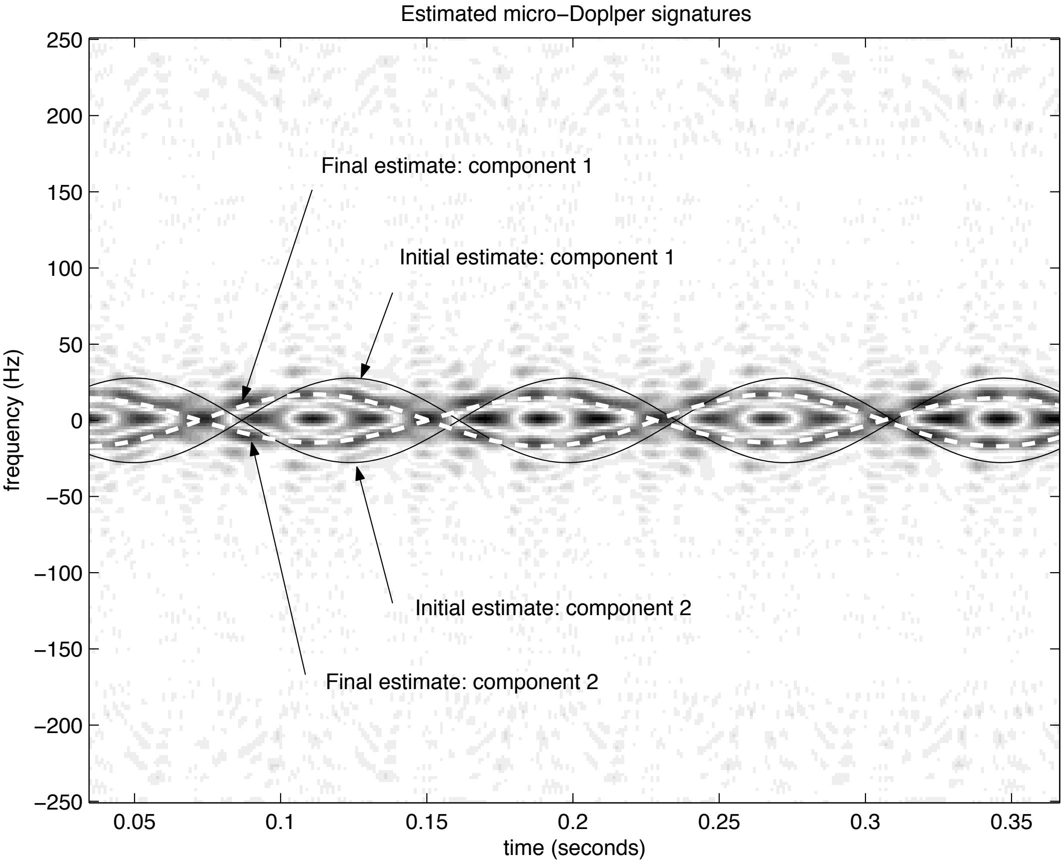


Figure 6.8: Estimated micro-Doppler signatures from the initial grid search with $L = 31$ (solid black curve) and optimization of the PWHT function with $L = 71$ (dashed white curve), overlaid on the PWVD of the data computed with $L = 71$.

6.A The micro-Doppler Cramér-Rao Bound

The CRB for $\boldsymbol{\eta}$ is the inverse of the Fischer information matrix (FIM), $\boldsymbol{J}$. For the complex Gaussian distribution, the ijth element of $\boldsymbol{J}$ is given by [106]:

$$J_{ij} = \mathsf{Tr}\left[\boldsymbol{\Gamma}^{-1}\frac{\partial\boldsymbol{\Gamma}}{\partial\eta_i}\boldsymbol{\Gamma}^{-1}\frac{\partial\boldsymbol{\Gamma}}{\partial\eta_j}\right] + 2\mathsf{Re}\left[\frac{\partial\boldsymbol{s}^H}{\partial\eta_i}\boldsymbol{\Gamma}^{-1}\frac{\partial\boldsymbol{s}}{\partial\eta_j}\right] \tag{6.20}$$

In the following we derive the quantities necessary to calculate the FIM using Equation (6.20). For compactness of notation in the following derivation, we define the following:

$$\begin{aligned}\boldsymbol{s}_d &= (e^{j\psi_d(0)}, \ldots, e^{j\psi_d(N-1)})' \\ \boldsymbol{s}_m &= (e^{j\psi_m(0)}, \ldots, e^{j\psi_m(N-1)})' \\ \boldsymbol{a} &= (a_1, \ldots, a_q)'\end{aligned}$$

The definitions from Equation (6.13) are also used here. The partial derivatives of the mean vector $\boldsymbol{s} = A_0 e^{j\theta_0}(\boldsymbol{s}_d \odot \boldsymbol{s}_m)$ with respect to the other parameters are given by:

$$\begin{aligned}\frac{\partial\boldsymbol{s}}{\partial B} &= j\boldsymbol{\varphi}_m \odot \boldsymbol{s} \\ \frac{\partial\boldsymbol{s}}{\partial\omega_m} &= j\boldsymbol{t}^{(1)} \odot \boldsymbol{\psi}_m \odot \boldsymbol{s} \\ \frac{\partial\boldsymbol{s}}{\partial\phi} &= j\boldsymbol{\psi}_m \odot \boldsymbol{s} \\ \frac{\partial\boldsymbol{s}}{\partial A_0} &= A_0^{-1}\boldsymbol{s} \\ \frac{\partial\boldsymbol{s}}{\partial\theta_0} &= j\boldsymbol{s} \\ \frac{\partial\boldsymbol{s}}{\partial a_i} &= j\boldsymbol{t}^{(i)} \odot \boldsymbol{s}\end{aligned}$$

To obtain the result Equation (6.12) we substitute the above into Equation (6.20), making use of the fact that $(\boldsymbol{w}\odot\boldsymbol{s})^H(\boldsymbol{z}\odot\boldsymbol{s}) = A_0^2 N\boldsymbol{w}^H\boldsymbol{z}$, for any length N vectors $\boldsymbol{w}$ and $\boldsymbol{z}$, where $\odot$ denotes the Hadamard (element-wise) matrix product.

6.B A sub-optimal estimation scheme

In this appendix, the "DP kernel" for a micro-Doppler signal is derived and some interesting properties are shown. We note that similar results were also obtained in [65] based on the higher-order ambiguity function (HAF). Based on these properties, a sub-optimal estimator for micro-Doppler parameters is derived.

6.B.1 The micro-Doppler DP kernel

We derive here the form of the DP kernel [93] for the micro-Doppler signal. The DP kernel can be considered a higher-order correlation function, whose Fourier transform is a higher-order ambiguity function. This kernel forms the basis of the polynomial phase transform [94], useful for estimation of polynomial phase coefficients. Let $c_x(t;k)$ denote the kth order DP kernel of a signal $x(t)$. Then $c_x(t;0) = x(t)$ and $c_x(t;k) = c_x(t+\tau;k-1)c_x^*(t;k-1)$ for $k \in \mathbb{Z}^+$. The DP kernel of the micro-Doppler signal $s_m(t)$ is given by

$$c_{s_m}(t;k) = e^{jB_k \sin(\omega_m t + \phi_k)}; \quad k \in \mathbb{Z}^+, \tag{6.21}$$

where $B_k = 2^k B \sin^k(\omega_m \tau/2)$ and $\phi_k = \phi + k(\omega_m \tau + \pi)/2$. We note that the DP kernel of $s_m(t)$ is a signal of exactly the same form, i.e. a sinusoidal phase term with the same micro-Doppler frequency ω_m.

Proof Let us denote the phase of $c_{s_m}(t;k)$ as $\psi_k(t) \equiv \angle c_{s_m}(t;k)$. From the definition of $\psi_m(t)$ in Equation (6.4) we see that $\psi_0(t) = B\sin(\omega_m t + \phi)$. Substituting this result into $\psi_k(t)$ for $k = 1$, we obtain

$$\begin{aligned}
\psi_1(t) &= \psi_0(t+\tau) - \psi_0(t) = \psi_m(t+\tau) - \psi_m(t) \\
&= B[\sin(\omega_m(t+\tau)+\phi) - \sin(\omega_m t + \phi)] \\
&= 2B\sin(\omega_m \tau/2)\cos(\omega_m t + \phi + \omega_m \tau/2) \\
&= B_1 \sin(\omega_m t + \phi_1)
\end{aligned}$$

where $B_1 = 2B\sin(\omega_m \tau/2)$ and $\phi_1 = \phi + (\omega_m \tau + \pi)/2$. Since $\psi_1(t)$ is of exactly the same form as $\psi_0(t)$, and $\psi_k(t) = \psi_{k-1}(t+\tau) - \psi_{k-1}(t)$ for $k \in \mathbb{Z}^+$, then we obtain $\psi_k(t) = B_k \sin(\omega_t + \phi_k)$ where

$$B_k = 2B_{k-1}\sin(\omega_m \tau/2) \quad \text{and} \quad \phi_k = \phi_{k-1} + (\omega_m \tau + \pi)/2, \tag{6.22}$$

for $k \in \mathbb{Z}^+$. Using the definition of B_1 and ϕ_1 in the above recursive formulas, we obtain $B_k = 2^k B \sin^k(\omega_m \tau/2)$ and $\phi_k = \phi + k(\omega_m \tau + \pi)/2$.

6.B.2 Polynomial phase approximation

Here we look at the approximation of the micro-Doppler phase as a polynomial. Using a Taylor series expansion, we can express the phase $\psi_m(t)$ in the vicinity of time t_0 as

$$\psi_m(t) = \sum_{k=0}^{\infty} \frac{\psi_m^{(k)}(t_0)}{k!}(t - t_0)^k \tag{6.23}$$

where $\psi_m^{(k)}(t)$ denotes the kth derivative of $\psi_m(t)$ with respect to t. Assuming the definition of $\psi_m(t)$ as given in Equation (6.4) and denoting the ith order coefficient of Equation (6.23) by b_i, we obtain the following coefficients (up to 2nd order, for $t_0 = 0$):

$$(b_0, b_1, b_2) = \left(B\sin(\phi), B\omega_m \cos(\phi), -\frac{1}{2} B\omega_m^2 \sin(\phi) \right).$$

Given the coefficients (b_0, b_1, b_2), one can find an inverse operation to retrieve the micro-Doppler parameters (B, ω_m, ϕ), but because the sin and cos functions are periodic and one may be expressed as a shifted version of the other, there is, in general, no unique inverse relationship. However, we may find a unique solution by restricting the parameter space so that $B > 0$, $\omega_m > 0$ and $\phi \in [0, 2\pi)$. In this case, we can obtain the micro-Doppler parameters from the phase coefficients successively:

$$\omega_m = \sqrt{\left| \frac{2b_2}{b_0} \right|} \tag{6.24}$$

$$B = \sqrt{b_0^2 + \left(\frac{b_1}{\omega_m} \right)^2} \tag{6.25}$$

$$\phi = \mathrm{sign}(b_0) \cdot \mathrm{acos}\left(\frac{b_1}{B\omega_m} \right) + 2\pi \cdot \mathrm{u}(-b_0) \tag{6.26}$$

where $\mathrm{u}(\cdot)$ denotes the unit step function which is one for a zero or positive argument and zero for a negative argument.

6.B.3 A Suboptimal Estimation Approach

Here we present a suboptimal approach to estimation of $\boldsymbol{\eta}_s$ based on the results presented in Sections 6.B.1 and 6.B.2. We assume the observation model of Equation (6.8) where the sampling period is known to be Δ seconds.

The proposed algorithm is detailed in Table 6.2 and summarised as follows: In the first step, the $(q+1)$th order DP kernel is computed to completely remove the Doppler phase term $\psi_d(t)$, leaving only the sinusoidal phase term as shown in Equation (6.21). In steps 2 and 3 a polynomial is fitted to the phase of the DP kernel and the amplitude, frequency and phase parameters are estimated based on the relationship derived in Equations (6.24)-(6.26). We then obtain the micro-Doppler parameters using the relationship in Equation (6.22). Finally, one may remove the estimated micro-Doppler term and estimate the other Doppler parameters using standard PPS algorithms.

1. Calculate the discrete $q+1$th order DP kernel of the data:

   $$z(n;\kappa) = \Pi_{k=1}^{q+1}\left(x^{\$k}(n+(k-1)\kappa)\right)^{\binom{q}{q+1-k}}; \quad n = 1,\ldots,N-\kappa(q+1).$$

   where

   $$x^{\$k}(n) = \begin{cases} x^*(n), & k \text{ even} \\ x(n), & k \text{ odd} \end{cases}$$

2. Fit a pth order polynomial to the phase of $z(n;\kappa)$, to obtain coefficients $(\hat{b}_0,\ldots,\hat{b}_p)$. This may be done using computationally efficient methods such as cyclic moments [85] or phase unwrapping followed by least squares [52].
   Note: the time reference should be adjusted here so that $t = 0$ occurs in the center of the interval.

3. Estimate ω_m, B_{q+1} and ϕ_{q+1} using

   $$\begin{aligned} \hat{\omega}_m &= \sqrt{\left|\frac{2\hat{b}_2}{\hat{b}_0}\right|} \\ \hat{B}_{q+1} &= \sqrt{\hat{b}_0^2 + \left(\frac{\hat{b}_1}{\hat{\omega}_m}\right)^2} \\ \hat{\phi}_{q+1} &= \operatorname{sign}(\hat{b}_0)\cdot \operatorname{acos}\left(\frac{\hat{b}_1}{\hat{B}_{q+1}\hat{\omega}_m}\right) + 2\pi\cdot \mathrm{u}(-\hat{b}_0) \end{aligned}$$

   and estimate B and ϕ using

   $$\begin{aligned} \hat{B} &= \hat{B}_{q+1}(2^{q+1}\sin^{q+1}(\hat{\omega}_m\kappa\Delta/2))^{-1} \\ \hat{\phi} &= \hat{\phi}_{q+1} - (q+1)(\hat{\omega}_m\kappa\Delta+\pi)/2 \end{aligned}$$

4. Demodulate the micro-Doppler component:

   $$\bar{x}(n) = x(n)\times\exp[-j\hat{B}\sin(\hat{\omega}_m n+\hat{\phi})]; \quad n = 1,\ldots,N$$

   and estimate $(a_0,\ldots,a_q)$ by applying a standard PPS estimation scheme.

Table 6.2: Estimation algorithm for the micro-Doppler signal model.

The application of the Taylor series expansion to the phase is done with respect to a certain time reference, t_0. When the phase is approximated by truncating the expansion, the error of the approximation increases as $|t - t_0|$ increases. It is therefore necessary to apply the polynomial fit to an appropriately chosen window of the phase, centered about the time t_0. We have found that using a window of one half of the micro-Doppler period and polynomial order 4 provides good results. We must first, however, have at least a rough estimate of ω_m in order to calculate the window length.

In the results given in the next section, we have used a simple correlation function approach to first estimate the micro-Doppler frequency. In step 2 of Table 6.2 we then average all available half-periods of the phase before estimating the polynomial coefficients. The half-period averaged phase term is calculated as follows:

1. $\psi(n) = \angle z(n;\kappa)$ for $n = 1, \ldots, N - \kappa(q+1)$.
2. $\psi(n) = K^{-1} \sum_{k=0}^{K} (-1)^k \psi(n + kM)$ for $n = 1, \ldots, M$ where $M = \lfloor \pi/(\hat{\omega}_m \Delta) \rfloor$ and $K = \lfloor (N - \kappa(q+1))/M \rfloor$.

It was found that correct selection of the lag κ to be used in computing the DP kernel is essential to achieving good performance. When using a small lag value (e.g. one sample) the amplitude of the micro-Doppler term is greatly reduced resulting in a lower 'phase SNR'. However, the value of κ should reduce the amplitude of the phase term just enough so that phase unwrapping is not required. The optimum value of κ is therefore signal dependent. From Equation (6.22) we see that the amplitude of the micro-Doppler term is reduced provided

$$\kappa < \frac{2}{\omega_m \Delta} \arcsin\left(2^{-(q+2)}\right). \tag{6.27}$$

Since B is generally greater than π, Equation (6.27) provides an upper bound on the value of κ to be used. One simple strategy to choose the appropriate value would be to start with this upper bound, and reduce κ until the number of phase jumps greater than a set threshold is reduced to an acceptable value. The upper-bound on κ can be estimated based on an the initial estimate of ω_m used in step 2.

6.B.4 Simulation results

In this section we present some simulation results illustrating the performance of the proposed estimation scheme. We use simulation parameters consistent with an X-band radar system tracking moving people. Results in [41, 42] clearly show the micro-Doppler effect is visible in these systems due to the swinging-arm action of a walking person. We also choose parameter values in line with real data presented in [42].

The X-band radar system has a carrier frequency of $f_c = 9.4$ GHz, a pulse repetition frequency $f_p = 800$ Hz, a dwell time of $N = 1024$ samples or 1.28 seconds and the speed of wave propagation is taken to be $c = 3 \times 10^8$ ms^{-1}. We assume a constant radial velocity of the walking person $v_r = 1.3$ ms^{-1} over the observation interval, with an arm swing rate of $f_a = 1.2$ Hz and swing amplitude $D = 0.23$ m. The signal model for this scenario is

$$s(t) = \exp[j(\omega_d t + B\sin(\omega_m + \phi) + \theta)]$$

where $\omega_m = 2\pi f_a$, $\omega_d = 4\pi f_c v_r/c$, $B = 4\pi f_c D/c$ and θ and ϕ are chosen arbitrarily on $[0, 2\pi)$.

The simulation results shown in Figure 6.9 show the result of phase unwrapping and averaging, followed by order 4 polynomial fitting. We can see that the phase is well approximated by the polynomial over a half-period of the micro-Doppler signature. We also show here the resulting phase estimate using the suboptimal procedure of Table 6.2, with an SNR of 10dB.

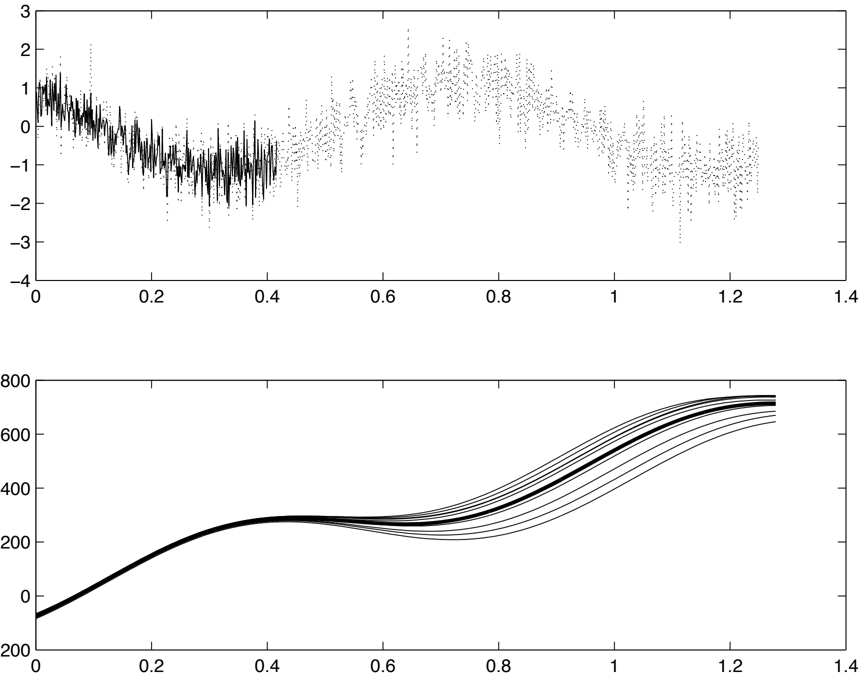


Figure 6.9: (Above) Phase after applying the DP kernel $(\cdots)$ and averaging over half-periods $(-)$. (Below) Estimated phase functions compared to true phase (thick line) using the proposed sub-optimal method.

The simulation results shown in Figure 6.10 compare the performance of the sub-optimal scheme with the CRB as the SNR is varied from -5 to 30 dB. We also plot the performance of the ML estimator given in Equation (6.11) when initialized with the sub-optimal estimate and the true parameter values. While the sub-optimal estimator is clearly biased and does not reach the CRB, is it sufficiently accurate for initialization of the ML estimator such that the CRB performance bound is achieved for high SNR. We also note that the RMSE is less than 1% of the true parameter value for all parameters, with SNR greater than 10 dB and for ω_m with SNR greater than 0 dB.

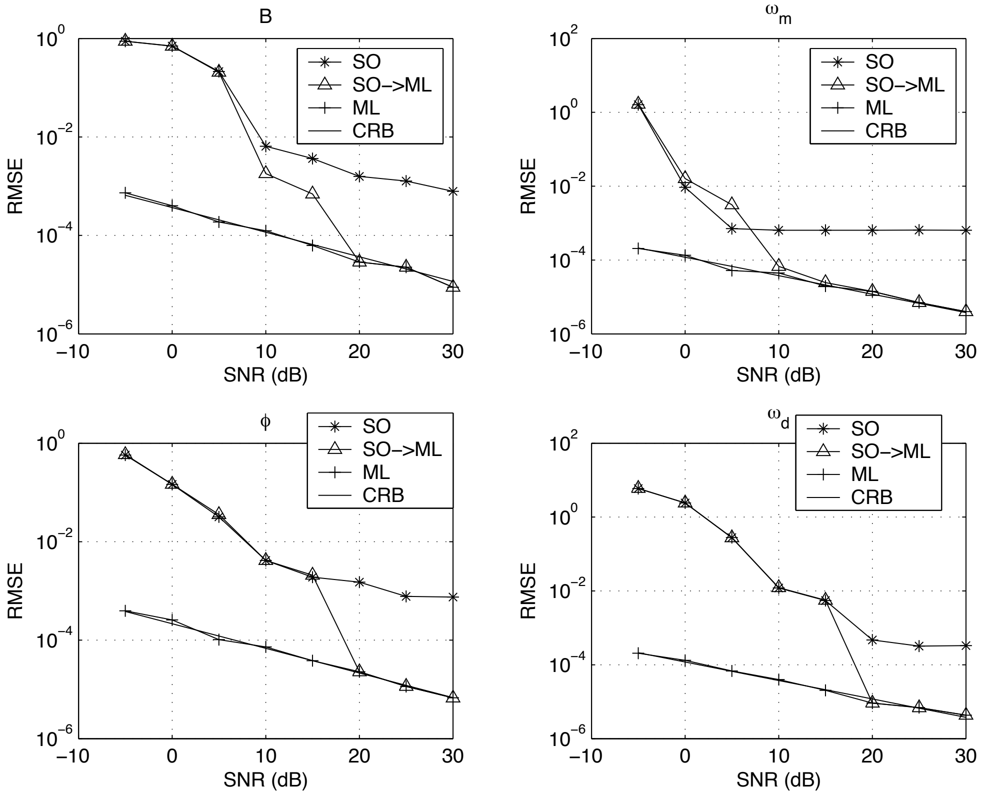


Figure 6.10: RMSE of sub-optimal and optimal estimators versus SNR.

Chapter 7

Near-Field Parameter Estimation

The array signal processing model was derived in Chapter 1 based on the far-field assumption. The estimation of far-field (DOA) parameters was reviewed in Section 2.2 and the corresponding time-frequency based approaches in Section 3.2.3. While the far-field DOA estimation problem has received considerable attention in the literature, certain applications arise in which the radiating or reflecting sources lie in the near-field of the sensor array. This chapter does not provide an in-depth study of the near-field parameter estimation problem. Rather, it is intended to supplement the main results of the this dissertation, by showing how the joint time-frequency approaches may be extended the near-field parameter estimation problem.

7.1 Introduction

The estimation of either far-field or near-field parameters have been mostly studied as separate problems. However, as discussed in [62], both near-field (NF) and far-field (FF) scatterers may be present in certain applications such as surface-wave radar. In such cases, estimation methods which assume only a given class (FF or NF) of scatterers are present mis-model the data and subsequently result in sub-optimal or erroneous estimation of the parameters of interest.

Characterization of NF scatterers in the presence of FF sources was considered in [62, 63], using a quadratic sensor-angle distribution (SAD). The work in [62] assumes that a NF scatterer is illuminated by a cooperative FF source of known DOA. Characterization of the NF scatterer requires one to first suppress the FF source using subspace projection techniques. Parameter estimation using the SAD, and the issue of aliasing, are discussed in [63]. It is noted that the SAD was formulated based on the assumption of a uniform linear array (ULA) geometry.

In this chapter, the estimation of NF parameters and a means of distinguishing between NF and FF sources is investigated, based on time-frequency analysis. The proposed approach does not require a particular array geometry, nor the DOA of FF sources to be estimated or known *a priori*. Estimation of NF parameters is performed using the same STFD matrices discussed in Section 3.2.

7.2 Signal Model

7.2.1 The near-field array response

If sources lie in the near-field of the sensor array, the response vector depends on the array geometry and the spatial location (both the range and direction) of the sources. As in Chapter 1, we assume that there are m sensors with locations $\{\boldsymbol{r}_k\}_{k=1}^{m}$ where the first (reference) sensor is located at the origin. It is assumed that the signal originates from a point in space described by the vector $\boldsymbol{q}$. The vector pointing from the kth sensor to the source location is denoted $\boldsymbol{d}_k \equiv \boldsymbol{q} - \boldsymbol{r}_k$, for $k = 1, \ldots, m$. The near-field array geometry is illustrated in Figure 7.1.

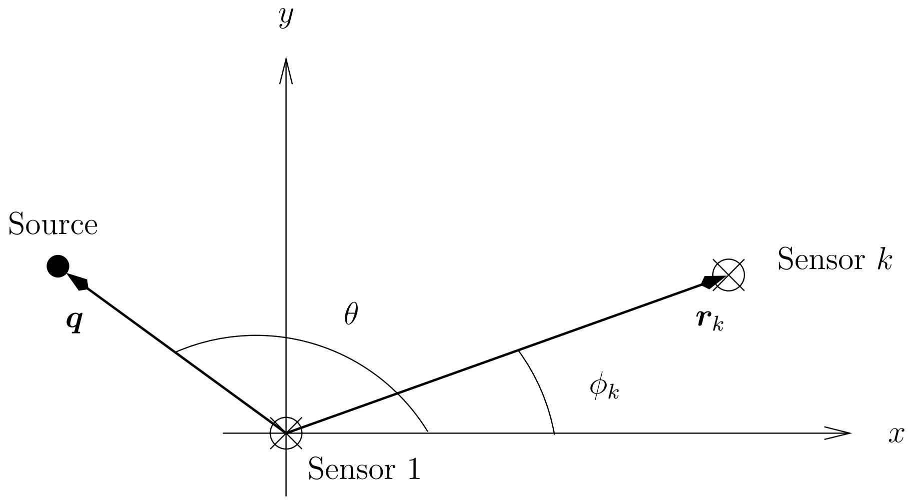


Figure 7.1: The near-field sensor array geometry.

Under the narrowband assumption, the source signal waveform at the origin can be expressed as $S(0,t) = s(t)e^{\mathrm{j}\omega_c t}$ where $s(t)$ is slowly time-varying with respect to the carrier term $e^{\mathrm{j}\omega_c t}$. The signal waveform at sensor k is then given by:

$$S(\boldsymbol{r}_k, t) = A_k s(t-\tau_k)e^{\mathrm{j}\omega_c (t-\tau_k)} \approx A_k s(t) e^{\mathrm{j}\omega_c (t-\tau_k)}, \qquad (7.1)$$

where A_k and τ_k respectively denote the amplitude and relative delay of the signal waveform at the kth sensor with respect to the reference sensor. Since the propagating signal amplitude is inversely proportional to the distance traveled, the relative amplitude is $A_k = \|\boldsymbol{d}_1\|/\|\boldsymbol{d}_k\|$. The relative delay is given by the difference in propagation times: $\tau_k = (\|\boldsymbol{d}_1\| - \|\boldsymbol{d}_k\|)/c$, where c is the speed of propagation.

For simplicity, it shall be assumed in the following that the sensor and source locations are co-planar. Equation (7.1) may be expressed as

$$S(\boldsymbol{r}_k, t) \approx \frac{d_1}{d_k} s(t) e^{\mathrm{j}\omega_c t} \mathrm{e}^{-\mathrm{j}\omega_c \frac{(d_1 - d_k)}{c}} = \left[\frac{d_1}{d_k} \mathrm{e}^{-\mathrm{j}\left(\frac{2\pi}{\lambda}\right)(d_1 - d_k)} \right] S(0, t), \tag{7.2}$$

where $d_k \triangleq \|\boldsymbol{d}_k\|$, for $k = 1 \ldots, m$, denotes the distance between the source and the kth sensor. Using the notation: $r_k \triangleq \|\boldsymbol{r}_k\|$ and $\phi_k \triangleq \angle \boldsymbol{r}_k$, for $k = 1 \ldots, m$, correspondingly $q \triangleq \|\boldsymbol{q}\|$ and $\theta \triangleq \angle \boldsymbol{q}$, the source-to-sensor distances may be calculated according to:

$$d_k = \sqrt{q^2 + r_k^2 - 2 q r_k \cos(\theta - \phi_k)}, \quad k = 1, \ldots, m. \tag{7.3}$$

Based on Equation (7.2) the NF array response (steering) vector for a source at location (q, θ) is defined as[1]

$$\boldsymbol{b}(q, \theta) = \left[1, \frac{d_1}{d_2} \mathrm{e}^{-\mathrm{j}\left(\frac{2\pi}{\lambda}\right)(d_1 - d_2)}, \ldots, \frac{d_1}{d_m} \mathrm{e}^{-\mathrm{j}\left(\frac{2\pi}{\lambda}\right)(d_1 - d_m)} \right]^{\mathrm{T}}. \tag{7.4}$$

where d_k depends on the source range, q and direction, θ, through the relationship given in Equation (7.3).

7.2.2 Baseband signal model

We consider an array of m sensors observing narrowband signals, of which J are in the NF and K in the FF of the array. The baseband signal model is given by

$$\boldsymbol{X}(t) = \sum_{k=1}^{J} \boldsymbol{b}(q_k, \theta_k) s_k(t) + \sum_{l=1}^{K} \boldsymbol{a}(\theta_l) z_l(t) + \boldsymbol{V}(t), \tag{7.5}$$

where $\boldsymbol{X}(t)$ is the vector of sensor outputs at time t, $\{s_k(t)\}$ are the NF source signals, $\{z_l(t)\}$ are the FF source signals and $\boldsymbol{V}(t)$ is an additive noise process. The vector $\boldsymbol{b}(q, \theta)$ denotes the array response to a NF source at range q and angle θ with respect to the reference sensor, as derived in the previous section. The

---

[1] As discussed in Chapter 1, it is assumed that the sensors are omni-directional with uniform gain.

vector $\boldsymbol{a}(\theta)$ denotes the array response to a FF source at angle θ. For convenience in the following, we may refer to the array response to the kth NF source as $\boldsymbol{b}_k$ and to the lth FF source as $\boldsymbol{a}_l$.

It is assumed that the signals $\{s_k(t)\}$ and $\{z_l(t)\}$ are deterministic signals with well-defined TF signatures and that noise is a complex white process of variance σ_V^2, satisfying Assumption 1. We model the sampled version of Equation (7.5) by the discrete-time random process $\boldsymbol{X}(n)$. *Given N observations, $\{\boldsymbol{x}(n)\}_{n=0}^{N-1}$, the problem is to estimate the spatial parameters, $\{(q_k, \theta_k)\}_{k=1}^{J}$, of the NF sources.*

7.3 Near-field parameter estimation

In this Section, we investigate the use of STFD matrices, for the purpose of NF parameter estimation. The STFD matrix of the sensor data is defined in Equation (3.22), whose elements are the auto- and cross-TFDs sensor data.

Let us assume the sources of interest have a well-defined TF structure, such as constant amplitude FM signals, whose TF signatures are distinct from one another. By averaging the STFD across the TF signature of a given source, one is able to isolate the signal energy from that component alone, and perform parameter estimation without the influence of the other signals. This property has been investigated for FF direction-finding in [124] and was summarized in Section 3.2.3.

Consider that the kth NF source has an instantaneous frequency (IF), denoted $\tilde{\omega}_k(t)$, for $k = 1, \ldots, J$. Calculating the averaged STFD for source k yields:

$$\boldsymbol{D}_k = \frac{1}{N} \sum_{n=0}^{N-1} \boldsymbol{D}_{\boldsymbol{xx}}(n, \tilde{\omega}_k(n)) \tag{7.6}$$

$$\approx \boldsymbol{b}(q_k, \theta_k) \left[\frac{1}{N} \sum_{n=0}^{N-1} D_{s_k s_k}(n, \tilde{\omega}_k(n)) \right] \boldsymbol{b}^{\mathrm{H}}(q_k, \theta_k) + \text{Noise term.} \tag{7.7}$$

Estimation of the NF parameters for source k may be achieved by applying traditional covariance based estimation methods to $\boldsymbol{D}_k$, assuming only one source is present. In the following, we shall make use of both the MUSIC estimator and the beamforming approach, which were reviewed for the FF case in Section 2.2.

The NF-MUSIC estimator is obtained from a subspace decomposition of the data covariance matrix in [76]. In this work, we apply the same approach to the averaged STFD matrix. An estimate of the range and bearing of the kth NF source is obtained according to:

$$(\hat{q}_k, \hat{\theta}_k) = \arg\max_{(q,\theta)} \frac{1}{\boldsymbol{b}^{\mathrm{H}}(q,\theta)\boldsymbol{U}_k\boldsymbol{U}_k^{\mathrm{H}}\boldsymbol{b}(q,\theta)}, \tag{7.8}$$

for $k = 1, \ldots, J$. In Equation (7.8), $\boldsymbol{U}_k$ denotes the noise subspace which is estimated from by taking the $m-1$ eigen-vectors of $\boldsymbol{D}_k$ corresponding to the $m-1$ smallest eigen-values.

In many applications, the sources are not well modelled by discrete points in space, which means that use of Equation (7.8) is not appropriate. Instead, the beamformer spectrum may be calculated over a range of parameter values to form an "image" of the NF characteristics of the scatterers (this shall be demonstrated in the following section using experimental data). The beamformer spectrum for the averaged STFD of source k is defined as:

$$P_k(q,\theta) \triangleq \frac{\boldsymbol{b}^{\mathrm{H}}(q,\theta)\boldsymbol{D}_k\boldsymbol{b}(q,\theta)}{\boldsymbol{b}^{\mathrm{H}}(q,\theta)\boldsymbol{b}(q,\theta)}; \quad k = 1, \ldots, J. \tag{7.9}$$

7.4 Near-field source discrimination

In order to apply Equation (7.6) for estimation, the IF of the NF sources must be known or estimated. In this section, we outline a means of discriminating the time-frequency signatures of NF sources from a mixture of NF and FF sources. The proposed approach is based on the TFDs of the sensor data.

We propose that by subtracting the average of sensor TFDs from the TFD computed at each sensor, one may remove the TF contributions of all FF sources. This is motivated by the fact the FF sources are received with the same power at each sensor, while the NF sources are received with varying powers due to the sphericity of the wavefront. Consider the following quantity:

$$\Delta_m(n,\omega) \triangleq D_{x_m x_m}(n,\omega) - \frac{1}{m}\sum_{l=1}^{m} D_{x_l x_l}(n,\omega), \tag{7.10}$$

for $m = 1, \ldots, m$. Substitution of the model from Equation (7.5) into (7.10) yields:

$$\begin{aligned}
\Delta_m(n,\omega) =& \sum_{u=1}^{J}\sum_{u=1}^{J}\left[b_{mu}b_{mv}^{*} - \frac{1}{m}\sum_{l=1}^{m} b_{lu}b_{lv}^{*}\right] D_{s_u s_v}(n,\omega) \\
&+ \sum_{u=1}^{K}\sum_{u=1}^{K}\left[a_{mu}a_{mv}^{*} - \frac{1}{m}\sum_{l=1}^{m} a_{lu}a_{lv}^{*}\right] D_{z_u z_v}(n,\omega) \\
&+ \sum_{u=1}^{J}\sum_{u=1}^{K}\left[b_{mu}a_{mv}^{*} - \frac{1}{m}\sum_{l=1}^{m} b_{lu}a_{lv}^{*}\right] D_{s_u z_v}(n,\omega) \\
&+ \sum_{u=1}^{K}\sum_{u=1}^{J}\left[a_{mu}b_{mv}^{*} - \frac{1}{m}\sum_{l=1}^{m} a_{lu}b_{lv}^{*}\right] D_{z_u s_v}(n,\omega) \\
&+ \text{Noise term,} \qquad (7.11)
\end{aligned}$$

for $m = 1, \ldots, m$, where b_{mv} and a_{mu} denote the mth element of $\boldsymbol{b}_v$ and $\boldsymbol{a}_u$ respectively.

We note that using a TFD with good cross-term suppression properties such as the Spectrogram, will make the terms in Equation (7.11) which involve $D_{s_u s_v}$ or $D_{z_u z_v}$, for $u \neq v$, and the terms $D_{z_u s_v}$ or $D_{s_u z_v}$ for all u, v, negligible. Assuming such a TFD is used in the computation of Equation (7.10), we may simplify Equation (7.11) to obtain:

$$\begin{aligned}
\Delta_m(n,\omega) =& \sum_{u=1}^{J}\left[|b_{mu}|^2 - \frac{1}{m}\sum_{l=1}^{m} |b_{lu}|^2\right] D_{s_u s_u}(n,\omega) \\
&+ \sum_{u=1}^{K}\left[|a_{mu}|^2 - \frac{1}{m}\sum_{l=1}^{m} |a_{lu}|^2\right] D_{z_u z_u}(n,\omega) \\
&+ \text{Noise term.} \qquad (7.12)
\end{aligned}$$

Under the mild assumption that each sensor of the array has the same gain, the magnitude of each element of the FF response vector is equal. The second term in Equation (7.12) therefore evaluates to zero and the quantity $\Delta_k(n,\omega)$ contains *only the TFDs of the near-field sources.*

As the weighting of NF source TFDs in Equation (7.12) may be negative, we propose the following "NF-TFD" for estimation of NF TF signatures:

$$B(n,\omega) \triangleq \left[\frac{1}{m}\sum_{m=1}^{m}\big(\Delta_m(n,\omega)\big)^2\right]^{\frac{1}{2}}. \qquad (7.13)$$

The TF signature of the NF sources may be estimated, e.g. by applying peak-finding or multi-component IF estimation techniques to $B(n,\omega)$. We note that, although the spectrogram is a good candidate to estimate the IF, as it does not exhibit cross-terms, calculation of the averaged STFD may be done using a higher-resolution TFD, such as the PWVD. The cross-terms present in the TFD are suppressed due to the averaging in Equation (7.6) and the improved TF resolution of the PWVD leads to more accurate estimation of the spatial parameters.

7.5 Simulation Results

The estimation accuracy of the proposed approach was evaluated via Monte Carlo simulations. In the experiment, two NF and two FF chirp signals were present. The NF sources had locations of $(q_1, \theta_1) = (0.5\lambda, 70°)$ and $(q_2, \theta_2) = (5\lambda, 69°)$ respectively, where λ denotes the carrier wavelength. The FF sources had directions $\theta_1 = 75°$ and $\theta_2 = 110°$. A uniform linear array structure of 6 elements with spacing $\lambda/2$ was used. All angles specified herein are w.r.t. the array endfire. All sources had the same power w.r.t. the noise and $N = 256$ observations were used. Calculation of the NF-TFD was based on a spectrogram with window length 51 samples.

The NF parameters were estimated using the NF-MUSIC algorithm. The RMSE of the estimates of range and direction of the first NF source are shown in Figure 7.2. For comparison, the estimation was performed using the STFD matrix when the TF signature of the source is known and when it was estimated using Equation (7.13). The signature was estimated by taking the location of the largest peak of $B(n, \omega)$ at each time-slice. In Figure 7.2, the accuracy of the MUSIC algorithm based on the covariance matrix and the CRB for estimation of a single NF source are also shown.

The results in Figure 7.2 indicated that STFD-based estimation is superior to covariance-based estimation at very low SNR. Unfortunately, the estimation of the NF TF signature is not successful at very low SNR where the best performance gain of the STFD-based estimation is achieved. We note, however, that in the range from 2 to 15 dB SNR the method which estimates the TF signature has higher accuracy than the covariance based method. We note that the application of better techniques for estimating the TF signature from $B(n, \omega)$, such as the Hough transform discussed in Chapter 5, will allow one to apply this approach at lower SNR.

7.6 Experimental Results

The proposed approaches for estimation of NF TF signatures and NF parameters have been applied to experimental data for validation of the theoretical ideas. The experimental system consists of a linear array of 16 antennas, located on the coast close to the land-sea boundary. Three targets are present in the experiment; two ships in the far-field of the array and one target in the near-field. All targets are transmitting linear FM signals of bandwidth 20 KHz and waveform repetition frequency 50 Hz, transmitted at a carrier frequency of 6.41 MHz.

The TF distribution of the signal received at the first sensor of the array and the NF-TFD of the sensor data are shown in Figures 7.3 and 7.4 respectively, which have been computed using the spectrogram with a rectangular window of length 61 samples. The NF-TFD clearly shows the TF signature of the NF source, while the TF signatures of the FF sources have been suppressed by over 15 dB.

Based on the NF-TFD shown in Figure 7.4, the IF of the NF source is estimated as the location of the largest peak for each time-slice. Using the estimated IF, the averaged STFD for the NF source is computed according to Equation (7.6), using the pseudo Wigner-Ville distribution (PWVD) with rectangular window of length 61. The NF beamformer spectrum is obtained according to Equation (7.9) and plotted in Figure 7.5. For comparison, the standard beamformer, obtained by substituting the matrix $\boldsymbol{D}_k$ with the sample covariance matrix in Equation (7.9), is plotted in Figure 7.6. We note that the use of TF processing to isolate the NF source allows one to more clearly observe the NF characteristics, without the influence of the FF sources.

7.7 Conclusions

The use of STFD matrices for NF parameter estimation has been investigated, and shown to provide improved accuracy with respect to covariance-based estimation. It was shown that by selectively averaging the TF signature of only NF sources, one may ignore other FF sources present in the data. A means of discriminating between the NF and FF TF signatures was also proposed and successfully applied to both simulated and experimental data.

7.8 Acknowledgment

I gratefully acknowledge the support of Dr. Gordon Frazer from the Australian Defence, Science and Technology Organization (DSTO), who provided me with the experimental data.

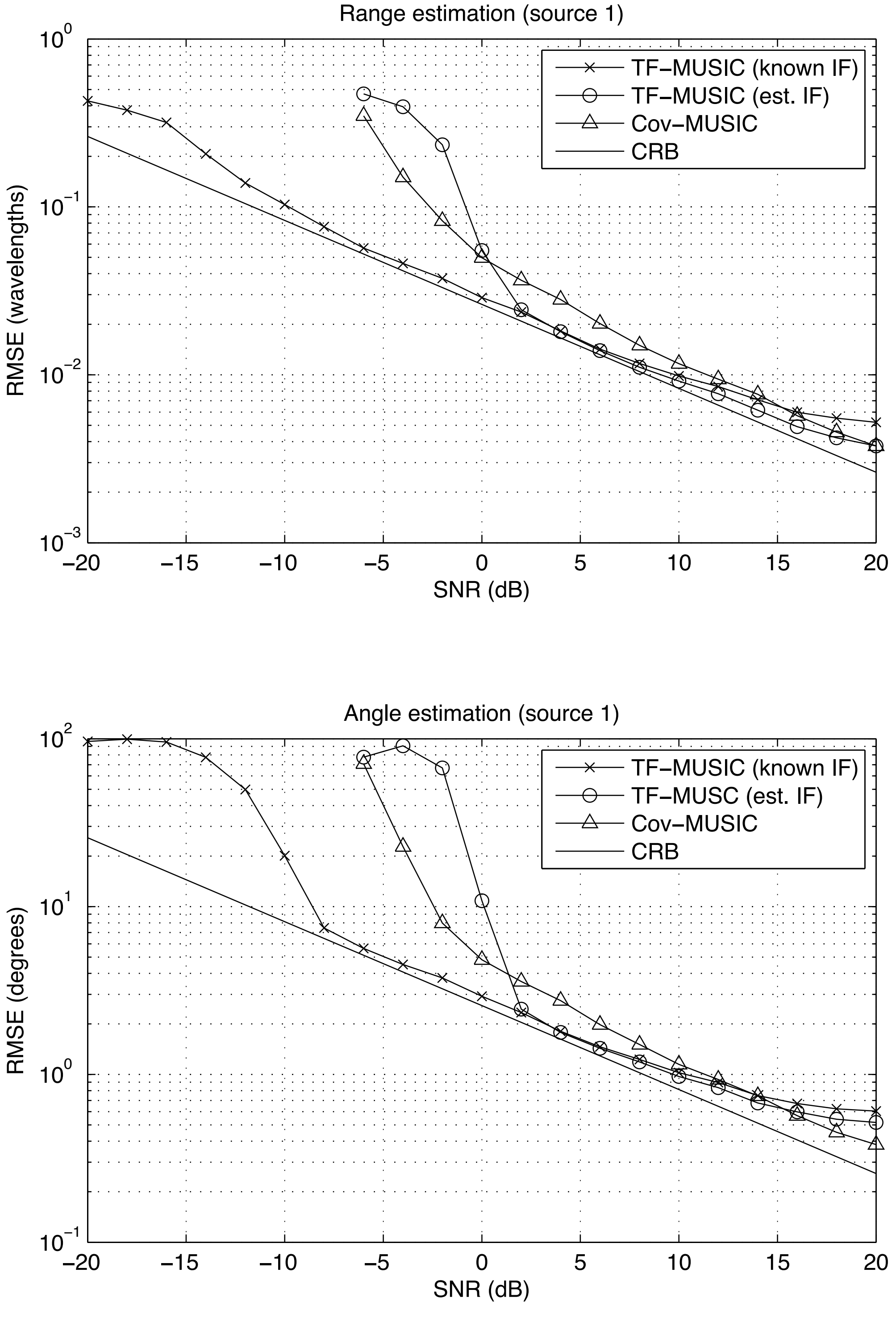


Figure 7.2: The RMSE of NF parameter estimation using the MUSIC algorithm, for estimation of the range (top) and direction (bottom) of the reference source.

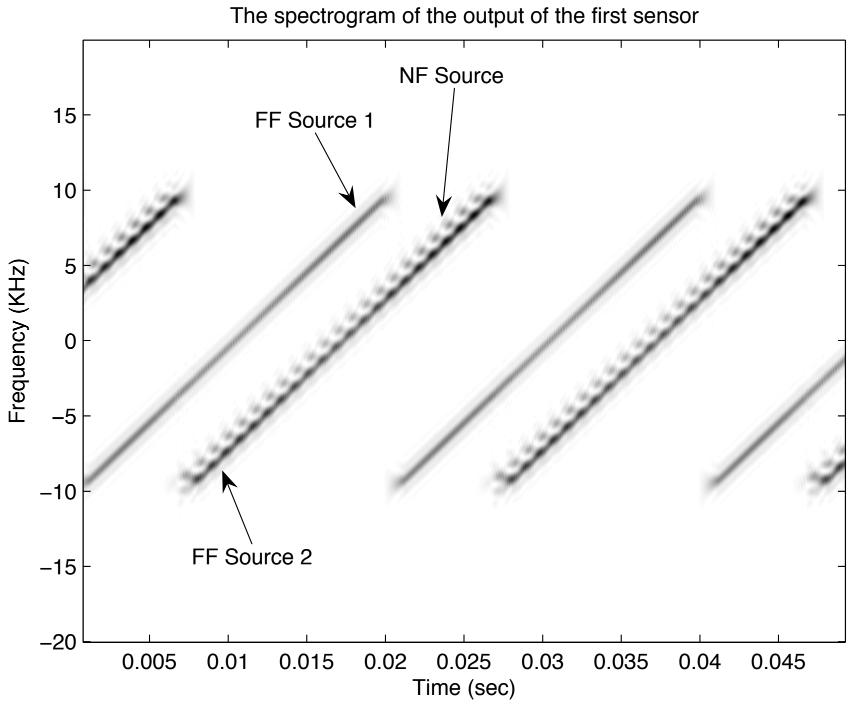


Figure 7.3: The TFD of the signal received at the first sensor of the experimental system.

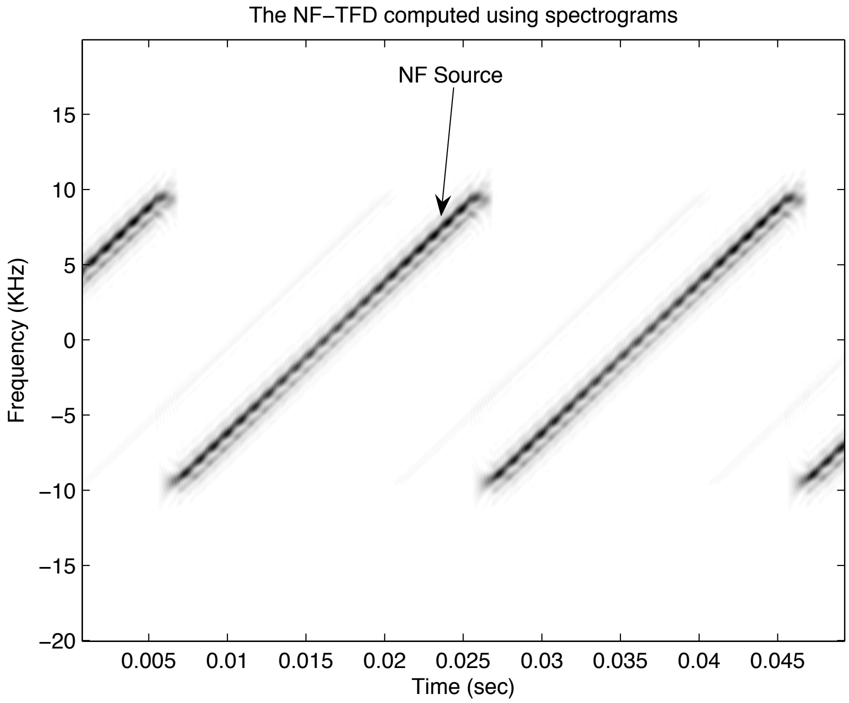


Figure 7.4: The NF-TFD of the experimental data, showing clearly the TF signature of the NF source.

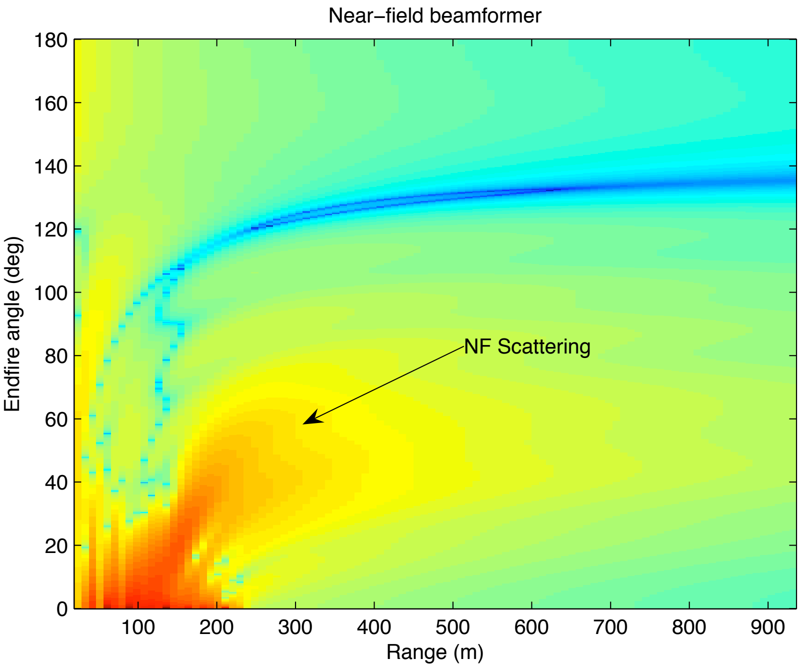


Figure 7.5: The NF beamformer obtained using the averaged STFD.

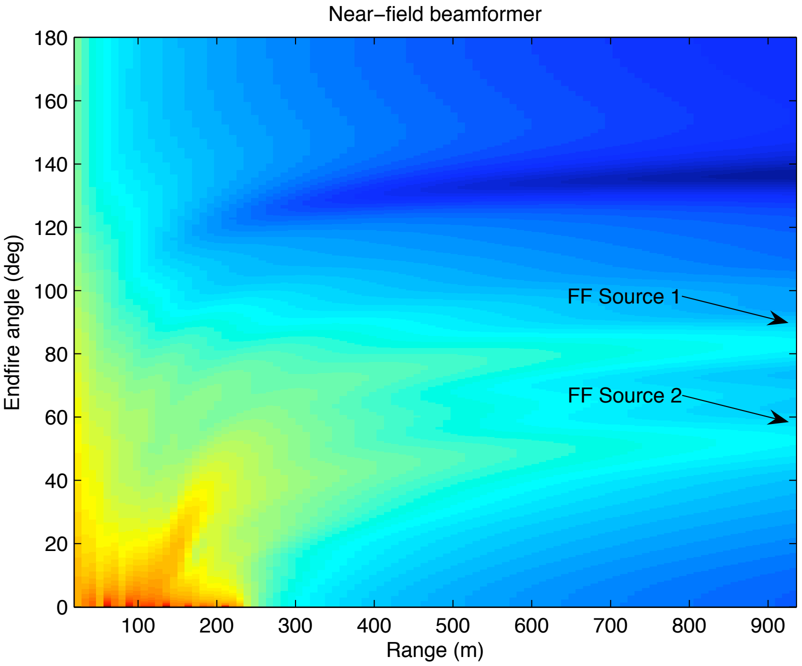


Figure 7.6: The NF beamformer obtained using the sample covariance matrix.

Chapter 8

Conclusions and Outlook

This chapter presents an overview of the conclusions drawn from the analysis and results of the preceding chapters. An outlook on the possible future directions for research in the field is presented in Section 8.2.

8.1 Conclusions

Time-frequency distributions are powerful tools for solving array signal processing problems, where nonstationarity or time-varying properties are characteristic of the underlying problem. The STFD formulation of the array processing model provides a natural means for exploiting both the spatial diversity and the TF localization properties of nonstationary sources impinging on a sensor array. There is, however, one caveat in the application of STFD methods: the TF localization of the sources should be known *a priori*, or must needs be estimated. The task of determining the TF localization properties of signals from noise-contaminated mixtures, has been composed as problems of 'TF point selection' and 'TF signature estimation' in this dissertation.

A 'point selection' scheme for automatically determining the TF locations at which STFD matrices exhibit underlying diagonal or off-diagonal structure was proposed, based on multiple hypothesis testing, in Chapter 4. The tendered method allows blind application of BSS based on the JD or JOD of STFD matrices. In contrast to other existing point selection schemes, there is no *ad hoc*, signal dependent threshold value to be chosen, rather, one decides on an acceptable probability of falsely selecting TF points. Further, it was found that the proposed approach could successfully determine TF points leading to favorable BSS performance, in cases of low SNR, closely spaced signals and closely overlapping TF signatures, where other methods fail.

Toward the goal of automatic TF signature estimation, a computationally attractive implementation of a time-frequency Hough transform (the PWHT) was offered in Chapter 5. A technique for reducing the number of required TF trajectories to be computed was proposed. It was demonstrated that one could effectively control the RoA of the estimator by varying the window length of the PWVD. An estimation scheme for both single and multicomponent signals was proposed based on this result. The application of the PWHT to nonlinear FM signals was also considered, and shown to be effective under the certain assumptions. Statistical analysis was performed in the case of additive white Gaussian noise. It was found that, in the case of linear FM signals there is clearly an optimal choice for the PWVD window length which yields estimation error close to the CRB.

The application of the proposed TF signature estimation to the problem of DOA estimation was also considered in Chapter 5. The considered approach was based on use of STFD matrices. Prior knowledge of the source TF signatures was not assumed, but estimated according to a parametric model using the TF Hough transform. Two important conclusions of this part of the work are as follows: Firstly, the use of the Hough transform provides performance close to the case of exactly known TF signatures at low SNR, where there is a significant gain in performance to be achieved by using STFD direction finding, over traditional methods. Secondly, the Hough transform allows DOA estimation for more sources than sensors, in an automatic way.

The particular problem of micro-Doppler signature estimation has also been studied in this dissertation. Due to rotational or vibrational movement of targets in a radar system, the received signal may exhibit a sinusoidal frequency modulation, known as micro-Doppler. The PWHT proposed in Chapter 5 has been investigated as a potential estimator for the sinusoidal FM parameters. It was found that the estimation of the micro-Doppler amplitude parameter is biased using the PWHT, and a bias correction scheme was proposed. The PWHT was applied to experimental radar data and demonstrated to successfully estimate the TF signatures of the micro-Doppler signal resulting from a rotating fan.

It is envisioned that the contributions made on the topic of TF signature estimation could prove useful in engineering applications such as radar target localization and identification.

As a final consideration, the use of STFD matrices for near-field parameter estimation was investigated in Chapter 7, and shown to provide improved accuracy with respect to existing methods. It was observed that by selectively averaging the TF signature of only NF sources, one may effectively ignore other FF sources present in the data. A means of discriminating between the NF and FF TF signatures was also proposed and verified by both simulation and the application to experimental data from a surface-wave radar system. The proposed approach could potentially be used to aid the analysis and characterization of NF scattering in sensor array systems.

8.2 Outlook

There are a number of interesting areas where the future research on array processing for nonstationary signals could be focused:

1. *Point selection*:

   - The test statistic for auto-term selection considered in this dissertation is based on a whitening transform, which can limit the achievable BSS performance. New test statistics should be studied, which do not depend on the whitening transform.
   - The test statistic proposed for cross-term selection requires the assumption of far-field sources and the sensors to have uniform gain. It is desirable to find other cross-term statistics which relax these constraints and allow for more general application of the techniques.

2. *TF signature estimation*:
   The time-frequency Hough transform proposed in this dissertation for TF signature estimation has a number of drawbacks: Firstly, one must know (at least approximately) a suitable parametric model for the signal IFs in a given application. If many parameters are required to sufficiently describe the IF, then estimation using the Hough transform becomes computationally inhibitive. Further, if there are mixtures of signals having different IF parametrizations, the application of the Hough transform is not straightforward. The development of *non-parametric TF signature estimation* methods for BSS and DOA estimation could potentially offer better flexibility and lower computational complexity.

3. *Use of polarization information*:
   An inherent assumption in the model used here, was that the sensors are delivering single-polarization information only. Methods which can exploit

multiple-polarizations should also be investigated. The particular relevance of polarized sensors to time-frequency based approaches should be examined.

4. *Stochastic models for source signals*:
   In this work, the source signals have been modeled as deterministic (unknown) waveforms. However, in particular applications, the underlying physical characteristics might motivate the use of stochastic signal models. The analysis of time-frequency based array processing methods and the extension of point selection and signature estimation techniques could be investigated under different stochastic models.

5. *Wideband and convolutive models*:
   Only narrowband array processing methods were considered in this dissertation. In applications such as sonar, telecommunications and speech signal processing, wideband signal models or convolutive mixture models may be more realistic in many scenarios. The extension of time-frequency based array processing methods to these cases would be of great theoretical and practical interest.

6. *Source number detection*:
   The problem of source number detection, or model order selection, has not been considered in this work. While there exist many methods for detecting the number of sources using a sensor array, these approaches do not, in general, exploit nonstationary signal properties. The extension of traditional methods or development of new approaches for detecting the number of *nonstationary* sources is an important problem.

Bibliography

[1] M. G. Amin A. Belouchrani, K. Abed-Meraim and A. M. Zoubir. Joint anti-diagonalization for blind source separation. In *Proceedings of the 26*$^{\mathrm{th}}$ *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2001.

[2] S. Amari. Neural learning in structured parameter spaces–natural Riemannian gradient. In *Proc. Neural Information Processing Systems-Natural and Synthetic (NIPS)*, December 1996.

[3] S. Amari, A. Cichocki, and H. H. Yang. A new learning algorithm for blind signal separation. *Advances in neural information processing systems*, 8:757–763, 1996.

[4] M. Amin, P. Zemany, P. Setlur, and F. Ahmad. Moving target localization for indoor imaging using dual frequency CW radars. In *IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, pages 367–371, July 2006.

[5] M. G. Amin. Spectal decomposition of time-frequency distribution kernels. *IEEE Transactions on Signal Processing*, 42(5):1156–1165, May 1994.

[6] M. G. Amin. Minimum variance time-frequency distribution kernels for signals in additive noise. *IEEE Transactions on Signal Processing*, 44(9):2352–2356, September 1996.

[7] M. G. Amin and Y. Zhang. Direction finding based on spatial time-frequency distribution matrices. *Digital Signal Processing*, 10(4):325–359, Oct 2000.

[8] J. Ängeby. Estimating parameters using the nonlinear instantaneous least squares approach. *IEEE Transactions on Signal Processing*, 48(10):2721–2732, Oct 2000.

[9] A. Barabell. Improving the resolution performance of eigen-structure based direction-finding algorithms. In *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Boston, MA, U.S.A, 1983.

[10] S. Barbarossa. Analysis of multicomponent LFM signals by a combined Wigner-Hough transform. *IEEE Transactions on Signal Processing*, 43(6):1511–1515, Jun 1995.

[11] S. Barbarossa and O. Lemoine. Analysis of nonlinear FM signals by pattern recognition of their time-frequency representation. *IEEE Signal Processing Letters*, 3(4):112–115, Apr 1996.

[12] S. Barbarossa and V. Petrone. Analysis of polynomial phase signals by the integrated generalized ambiguity function. *IEEE Transactions on Signal Processing*, 45(2):316–327, Feb 1997.

[13] S. Barbarossa, A. Scaglione, and G. Giannakis. Product higher-order ambiguity function for multicomponent polynomial-phase signal modeling. *IEEE Transactions on Signal Processing*, 46(3):691–708, Mar 1998.

[14] B. Barkat and B. Boashash. Reply to "comments on 'a high-resolution quadratic time-frequency distribution for multicomponent signals analysis'". *IEEE Transactions on Signal Processing*, 52(1):290–291, Jan 2004.

[15] A. J. Bell and T. J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural computation*, 7(6):1004–1034, 1995.

[16] M. Bell and R. Grubbs. JEM modeling and measurement for radar target identification. *IEEE Transactions on Aerospace and Electronic Systems*, 29(1):73–87, January 1993.

[17] A. Belouchrani, K. Abed-Meraim, J.-F.Cardoso, and E. Moulines. A blind source separation technique using second-order statistics. *IEEE Transactions on Signal Processing*, 45(2):434–443, February 1997.

[18] A. Belouchrani, K. Adeb-Meraim, M. G. Amin, and A. M. Zoubir. Blind source separation of nonstationary signals. *IEEE Signal Processing Letters*, 11(7):605–608, Jul 2004.

[19] A. Belouchrani and M. Amin. Source separation based on the diagonalization of a combined set of spatial time-frequency distributions. *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 1997. Germany.

[20] A. Belouchrani and M. Amin. Blind source separation based on time-frequency representations. *IEEE Transactions on Signal Processing*, 46(11):2888–2897, November 1998.

[21] A. Belouchrani, M. Amin, and K. Abed-Meraim. Direction finding in correlated noise fields based on joint block-diagonalization of spatio-temporal correlation matrices. *IEEE Signal Processing Letters*, 4(9):266–268, September 1997.

[22] A. Belouchrani and M. G. Amin. Time-frequency music. *IEEE Signal Processing Letters*, 1999.

[23] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc. B*, 57(1):289–300, 1995.

[24] Y. Benjamini and Y. Hochberg. Multiple hypotheses testing with weights. *Scandinavian Journal of Statistics*, 24(3):407–418, 1997.

[25] B. Boashash and A. Reilly. *Time-Frequency Signals Analysis - Methods and Applications*, chapter Algotihms for time-frequency signal analysis. Longman Cheshire Pty, 1995.

[26] F. F. Böhme. Estimation of source parameters by maximum likelihood and nonlinear regression. In *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1984.

[27] J. F. Böhme. Estimation of spectral parameters of correlated signals in wavefields. *Signal Processing*, 11(4):329–327, December 1986.

[28] R. Brcich, A. M. Zoubir, and P. Pelin. Detection of sources using bootstrap techniques. *IEEE Transactions on Signal Processing*, 50(2):206–215, February 2002.

[29] Y. Bresler and A. Macovski. Exact maximum likelihood parameter estimation of superimposed exponential signals in noise. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 34(10):1081–1089, October 1986.

[30] K. Buckley and X.-L. Xu. Spatial spectrum estimation in a location sector. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 38(11):1842–1852, November 1990.

[31] J. P. Burg. The relationship between maximum entropy spectra and maximum likelihood spectra. *Geophysics*, 37:375–376, 1972.

[32] J. A. Cadzow. Multiple source location - the signal subspace approach. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 38(7):1110–1125, July 1990.

[33] J.-F. Cardoso. On the performance of source separation algorithms. In *Proc. European signal processing conference (EUSIPCO)*, pages 776–779, Edinburgh, 1994.

[34] J.-F. Cardoso. Infomax and maximum likelihood for blind source separation. *IEEE Signal Processing Letters*, 4(4):112–114, April 1997.

[35] J.-F. Cardoso. Statistical principles of source separation. In *Proc. IFAC IFAC symposium on system identification (SYSID)*, volume 3, pages 1837–1844, 1997.

[36] J.-F. Cardoso. Higher-order contrasts for independent component analysis. *Neural computation*, 11(1):157–192, January 1999.

[37] J.-F. Cardoso and B. Laheld. Equivariant adaptive source separation. *IEEE Transactions on Signal Processing*, 44(12):3017–3030, December 1996.

[38] J.-F. Cardoso and A. Souloumiac. Blind beamforming for non-Gaussian signals. *Proceedings of the IEEE*, 140(6):362–370, December 1993.

[39] V. C. Chen. Analysis of radar micro-Doppler signature with time-frequency transform. In *Proc. of the 10th IEEE Workshop onStatistical Signal and Array Processing (SSAP)*, pages 463–465, August 2000.

[40] V. C. Chen, F. Li, S.-S. Ho, and H. Wechsler. Analysis of micro-Doppler signatures. *IEE Proc. on Radar, Sonar and Navigation*, 150(4):271–276, August 2003.

[41] V. C. Chen, F. Li, S.-S. Ho, and H. Wechsler. Micro-Doppler effect in radar: phenomenon, model, and simulation study. *IEEE Transactions on Aerospace and Electronic Systems*, 42(1):2–21, January 2006.

[42] V. C. Chen, R. Lipps, and M. Bottoms. Advanced synthetic aperture radar imaging and feature analysis. In *Proc. of the International Radar Conference*, pages 22–29, September 2003.

[43] L. Cirillo and M. Amin. Auto-term detection using time-frequency array processing. In *Proceedings of the 28*$^{\text{th}}$ *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Hong Kong, China, April 2003.

[44] L. Cirillo, A. Zoubir, and M. Amin. Selection of auto- and cross-terms for blind non-stationary source separation. In *Proceedings of the 2*$^{\text{nd}}$ *IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, Cairo, Egypt, December 2001.

[45] L. Cirillo, A. Zoubir, N. Ma, and M. Amin. Automatic classification of auto- and cross-terms of time-frequency distributions in antenna arrays. In *Proceedings of the 27*$^{\text{th}}$ *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Orlando, FL USA, May 2002.

[46] T. A. C. M. Claasen and W. F. G. Mecklenbrauker. The Wigner distribution – a tool for time-frequency signal analysis, part III: relations with other time-frequency signal transformations. *Philips J. Res.*, 35:372–389, 1980.

[47] L. Cohen. Generalized phase-space distribution functions. *J. math. phys.*, 7:781–786, 1966.

[48] L. Cohen. *Time-Frequency Analysis.* Prentice Hall, 1995.

[49] P. Common. Independent component analysis, a new concept? *Signal Processing*, 36(3):287–314, April 1994.

[50] G. S. Cunningham and W. J. Williams. Fast implementations of generalized discrete time-frequency distributions. *IEEE Transactions on Signal Processing*, 42(6):1496–1508, June 1994.

[51] G. S. Cunningham and W. J. Williams. Kernel decomposition of time-frequency distributions. *IEEE Transactions on Signal Processing*, 42(6):1425–1442, June 1994.

[52] P. M. Djurić and S. M. Kay. Parameter estimation of chirp signals. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 38(12):2118–2126, December 1990.

[53] B. Efron and R. Tibshirani. *An Introduction to the Bootstrap.* Chapman and Hall, 1993.

[54] N. Thirion-Moreau E.M. Fadaili and E. Moreau. Combined non-orthogonal joint zero-diagonalization and joint diagonalization for source separation. In *Proc. of IEEE Workshop on Statistical Signal Processing (SSP)*, pages 1120–5, Bordeaux, France, July 2005.

[55] N. Thirion-Moreau E.M. Fadaili and E. Moreau. Non-orthogonal zero-diagonalization for source separation based on time-frequency representation. In *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, volume IV, Philadelphia, PA USA, March 2005.

[56] D. Farina, C. Févotte, C. Doncarli, and R. Merletti. Blind separation of instantaneous mixtures of nonstationary surface myoelectric signals. *IEEE Transactions on Biomedical Engineering*, 51(9):1555–1567, Sep 2004.

[57] D. Farrier and L. Prosper. A signal subspace beamformer. In *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Albuquerque, NM, U.S.A, 1990.

[58] D. Farrier and L. Prosper. Threshold extension based on a new paradigm for MUSIC-type estimation. In *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Albuquerque, NM, U.S.A, 1990.

[59] C. Févotte and C. Doncarli. Two contributions to blind source separation using time-frequency distributions. *IEEE Signal Processing Letters*, 11(3):386–389, Mar 2004.

[60] R. Fletcher and M. Powell. A rapidly convergent descent method for minimization. *Computer J.*, 6:163–168, 1963.

[61] J. C. Fort. Stability of the JH source separation algorithm. *Traitement du Signal*, 8(1):35–42, 1991.

[62] G. J. Frazer and M. G. Amin. Characterization of near-field scattering using quadratic sensor-angle distributions. In *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, volume III, pages 2825–2828, Orlando, FL, USA, May 2002.

[63] G. J. Frazer and M. G. Amin. Near-field scatter measurements using quadratic sensor-angle distributions. In *Proc. of IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, pages 72–76, Rosslyn, VA, USA, August 2002.

[64] M. Gaeta and J. L. Lacoume. Source separation without prior knowledge: The maximum likelihood solution. In *Proc. European signal processing conference (EUSIPCO)*, pages 621–624, 1990.

[65] F. Gini and G. B. Giannakis. Hybrid FM-polynomial phase signal modeling: parameter estimation and Cramér-Rao bounds. *IEEE Transactions on Signal Processing*, 47(2):363–377, February 1999.

[66] L. Giulieri, H. Ghennioui, N. Thirion-Moreau, and E. Moreau. Nonorthogonal joint diagonalization of spatial quadratic time-frequency matrices for source separation. *IEEE Signal Processing Letters*, 12(5):415–418, May 2005.

[67] L. Giulieri, N. Thirion-Moreau, and P.-Y. Arques. Blind sources separation based on bilinear time-frequency representations: A performance analysis. In *Proceedings of the 27th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Orlando, FL USA, May 2002.

[68] J. E. Gray and S. R. Addison. Effect of nonuniform target motion on radar backscattered waveforms. *IEE Proc. on Radar, Sonar and Navigation*, 150(4):262–270, August 2003.

[69] J. Hérault, C. Jutten, and B. Ans. Détection de grandeurs primitives dans un message composite par une architecture de calcul neuromimétique en apprentissage non supervisé. In *Proc. Xème collouqe GRETSI*, pages 1017–1022, Nice, France, May 1985.

[70] Y. Hochberg. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*, 75(4):800–802, 1988.

[71] Y. Hochberg and A. C. Tamhane. *Multiple Comparison Procedures.* John Wiley & Sons, 1987.

[72] S. Holm. A simple sequentially rejective multiple test procedure. *Scand J Statist*, 6:65–70, 1979.

[73] G. Hommel. A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika*, 75(2):383–6, 1988.

[74] G. Hommel. A comparison of two modified Bonferroni procedures. *Biometrika*, 76(3):624–5, 1989.

[75] S.-R. Huang, R. Lerner, and K. Parker. On estimating the amplitude of harmonic vibration from the Doppler spectrum of reflected signals. *Journal of the Acoustical Society of America*, 88(6):2702–2712, December 1990.

[76] Y.-D. Huang and M. Barkat. Near-field source localization by passive sensor array. *IEEE Transactions on Antennas and Propagation*, 39(7):968–975, July 1991.

[77] Z. M. Hussain. Comments on "a high-resolution quadratic time-frequency distribution for multicomponent signals analysis". *IEEE Transactions on Signal Processing*, 52(1):287–289, Jan 2004.

[78] A. G. Jaffer. Maximum likelihood direction finding of stochastic sources: a separable solution. In *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, New York, New York, U.S.A., April 1988.

[79] J. Jeong and W. Williams. Kernel design for reduced interference distributions. *IEEESP*, 40(2):402–411, February 1992.

[80] C. Jutten and J. Hérault. Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 24:1–10, 1991.

[81] H. Krim and M. Viberg. Two decades of array signal processing research. *IEEE Signal Processing Magazine*, pages 67–94, July 1996.

[82] E. L. Lehmann. *Testing Statistical Hypotheses.* John Wiley & Sons, second edition, 1986.

[83] E. L. Lehmann and G. Casella. *Theory of point estimation.* Springer-Verlag New York, In., second edition, 1998.

[84] L. Molgedey and H. G. Schuster. Separation of independent signals using time-delayed correlations. *Physics Reviews Letters*, 72(23):3634–3637, 1994.

[85] M. R. Morelande and A. M. Zoubir. On the performance of cyclic moments-based parameter estimators of amplitude modulated polynomial phase signals. *IEEE Transactions on Signal Processing*, 50(3):590–606, March 2002.

[86] E. Moulines, J.-F. Cardoso, and E. Gassiat. Maximum likelihood for blind separation and deconvolution of noisy signals using mixture models. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3617–3620, Munich, Germany, April 1997.

[87] W. Mu, M. G.Amin, and Y. Zhang. Bilinear signal synthesis in array processing. *IEEE Transactions on Signal Processing*, 51(1):90–100, Jan 2003.

[88] J. Nelder and R. Mead. A simplex method for function minimization. *Computer J.*, 7:308–313, 1965.

[89] P. Oliveria and V. Barroso. Sequential extraction of components of multi-component PPS signals. In *Proc. ICASSP*, Phoenix, USA, Mar 1999.

[90] B. Ottersten, M. Viberg, P. Stoica, and A. Nehorai. Analysis of subspace fitting and ML techniques for parameter estimation from sensor array data. *IEEE Transactions on Signal Processing*, 40(3):590–600, March 1992.

[91] B. Ottersten, M. Viberg, P. Stoica, and A. Nehorai. *Radar array processing*, chapter Exact and large sample ML techniques for parameter estimation and detection in array processing, pages 99–151. Springer-Verlag Berlin, Inc., 1993.

[92] B. A. Pearlmutter and L. C. Parra. A context-sensitive generalization of ICA. In *Proc. International conference on neural information processing*, Hong Kong, 1996.

[93] S. Peleg and B. Friedlander. The discrete polynomial-phase transform. *IEEE Transactions on Signal Processing*, 43(8):1901–1914, August 1995.

[94] S. Peleg and B. Porat. Estimation and classification of signals with polynomial phase. *IEEE Transactions on Information Theory*, 37:422–430, 1991.

[95] D.-T. Pham. Blind separation of instantaneous mixtures of sources via an independent component analysis. *IEEE Transactions on Signal Processing*, 44(11):2768–2779, November 1996.

[96] D.-T. Pham and P. Garat. Blind separation of independent sources through a quasi-maximum likelihood approach. *IEEE Transactions on Signal Processing*, 45(7):1712–1725, July 1997.

[97] A. W. Rihaczek. *Principles of high-resolution radar.* Artech House Publishers, 1996.

[98] L. Scharf. *Statistical signal processing: detection, estimation and time-series analysis.* Addison-Wesley, Reading, MA, 1992.

[99] R. Schmidt. Multiple emitter location and signal parameter estimation. *IEEE Transactions on Antennas and Propagation*, 34(3):276–280, March 1986.

[100] P. Setlur, M. Amin, F. Ahmad, and T. Thayaparan. Indoor imaging of targets enduring simple harmonic motion using Doppler radars. In *Proceedings of the International Symposium on Signal Processing and Information Technology (ISSPIT)*, pages 141–146, Athens, Greece, December 2005.

[101] M. I. Skolnik. *Introduction to RADAR systems.* McGraw-Hill, third edition, 2001.

[102] T. Sparr and B. Krane. Micro-Doppler analysis of vibrating targets in SAR. *IEE Proc. on Radar, Sonar and Navigation*, 150(4):277–283, August 2003.

[103] L. Stankovic. An analysis of noise in time-frequency distributions. *IEEE Signal Processing Letters*, pages 286–289, Sep 2002.

[104] D. Starer and A. Nehorai. Newton algorithms for conditional and unconditional maximum likelihood estimation of the parameters of exponential signals in noise. *IEEE Transactions on Signal Processing*, 40(6):1520–1534, June 1992.

[105] P. Stoica and A. Nehorai. MUSIC, maximum likelihood, and cramer-rao bound. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(5):720–741, May 1989.

[106] P. Stoica and A. Nehorai. Performance study of conditional and unconditional direction-of-arrival estimation. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 38(10):1783–1795, October 1990.

[107] P. Stoica and A. Nehorai. Performance study of conditional and unconditional direction-of-arrival estimation. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 38(10):1783–95, Oct 1990.

[108] P. Stoica and K. Sharman. Maximum likelihood methods for direction-of-arrival estimation. *IEEE Transactions on Signal Processing*, 38(7):1132–1143, July 1990.

[109] P. Stoica and K. Sharman. A novel eigenanalysis method for direction estimation. *Proc. of the IEE*, F:19–26, February 1990.

[110] K. Taylor. Absolute calibration of microphone by a laser Doppler. *Journal of the Acoustical Society of America*, 70:939–945, 1981.

[111] L. Tong, R. Liu, V. Soon, and Y. Huang. Indeterminacy and identifiability of blind identification. *IEEE Transactions on Circuits and Systems*, 21(5):499–509, May 1991.

[112] H. L. Van Trees. *Optimum array processing.* John Wiley & Sons, Inc., 2002.

[113] T. Venkatesan and M. G. Amin. Time-frequency distribution kernel design over a discrete powers-of-two space. *IEEE Signal Processing Letters*, 3(12):305–306, December 1996.

[114] M. Viberg, B. Ottersten, and T. Kailath. Detection and estimation in sensor arrays using weight subspace fitting. *IEEE Transactions on Signal Processing*, 39(11):2436–2449, November 1991.

[115] M. Viberg and A. L. Swindlehurst. A Baysian approach to auto-calibration for parametric array signal processing. *IEEE Transactions on Signal Processing*, 42(12):3495–3507, December 1994.

[116] M. Wax. Detection and localization of multiple sources in noise with unknown covariance. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 40(1):245–249, January 1992.

[117] J. E. Wilbur and R. J. McDonald. Nonlinear analysis of cyclically correlated spectral spreading in modulated signals. *Journal of the Acoustical Society of America*, 92(1):219–230, July 1992.

[118] M. Wong and Q. Jin. Estimation of the time-varying frequency of a signal: the Cramer-Rao bound and the application of Wigner distribution. *IEEE*

Transactions on Acoustics, Speech, and Signal Processing, 38(3):519–536, March 1990.

[119] J. C. Wood and D. T. Barry. Radon transform of time-frequency distributions for analysis of multicomponent signals. *IEEE Transactions on Signal Processing*, 42(11):3166–3177, Nov 1994.

[120] Y. Yamakoshi, J. Sato, and T. Sato. Ultrasonic imaging of internal vibration of soft tissue under forced vibration. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 37(2):45–53, March 1990.

[121] A. Yeredor. Non-orthogonal joint diagonalization in the least squares sense with application in blind source separation. *IEEE Transactions on Signal Processing*, 50(7):1545–1553, July 2002.

[122] Y. Zhang and M. G. Amin. Spatial averaging of time-frequency distributions for signal recovery in uniform linear arrays. *IEEE Transactions on Signal Processing*, 48(10):2892–902, October 2000.

[123] Y. Zhang, W. Mu, and M. G. Amin. Time-frequency maximum likelihood methods for direction finding. *Journal of the Franklin Institute*, 337(4):483–497, July 2000.

[124] Y. Zhang, W. Mu, and M. G. Amin. Subspace analysis of spatial time-frequency distribution matrices. *IEEE Transactions on Signal Processing*, 49(4):747–759, Apr 2001.

Curriculum Vitae

PERSONAL DETAILS

Name: Luke Aaron CIRILLO
Date of birth: 17 July, 1979
Place of birth: Perth, Australia
Family status: Married, a daughter

EDUCATION

1997-2000	Bachelor of Electrical Engineering (Electronics and Communications), Curtin University of Technology, Perth, Australia
1994-1996	Trinity College, Perth, Australia
1992-1993	Keanan College, Manjimup, Australia
1988-1991	St. Joseph's Catholic Primary, Bunbury, Australia
1985-1987	Mt. Baker Primary School, Mt. Baker, Australia

WORK EXPERIENCE

03/2003-	Research Associate, Darmstadt University of Technology, Darmstadt, Germany
03/2000-02/2003	Teaching and research assistant, Curtin University of Technology, Perth, Australia
12/1999-02/2000	Research Engineer, Australian Telecommunications Co-operative Research Center, Perth, Australia
12/1998-02/1999	Software developer, IMAGO Multimedia Centre, Perth, Australia